

# Van vakgericht naar competentiegericht statistiekonderwijs

Citation for published version (APA):

Van Buuren, H. (2008). *Van vakgericht naar competentiegericht statistiekonderwijs: Een interventiestudie in een opleiding psychologie*. [, Open Universiteit]. Open Universiteit.

## Document status and date:

Published: 13/06/2008

## Document Version:

Peer reviewed version

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

## Take down policy

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05 May. 2023



**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



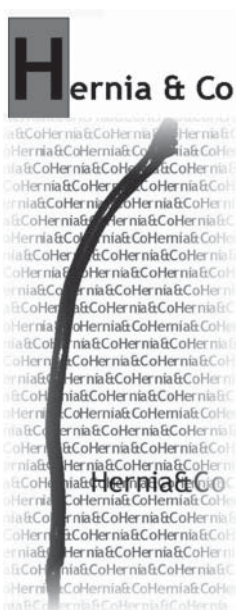


# **Van vakgericht naar competentiegericht statistiekonderwijs**

*Een interventiestudie in een opleiding psychologie*



Met bijdragen van:



*Scholing en begeleiding  
voor basisonderwijs*

Nel van Woerkum Caumerdalgestraat 14  
6416 GD Heerlen tel. 045-5713296

Omslagontwerp en lay-out: Annette Bouwels en Sandra Daems  
Drukwerk: Schrijen Lippertz Huntjens, Voerendaal

© 2008 J.A. van Buuren, Heerlen

ISBN 978-90-358-0817-1

# **Van vakgericht naar competentiegericht statistiekonderwijs**

*Een interventiestudie in een opleiding psychologie*

**From subject-oriented to competence-based statistics education  
An intervention study in a school of psychology**

PROEFSCHRIFT

ter verkrijging van de graad van doctor  
aan de Open Universiteit Nederland  
op gezag van de rector magnificus  
prof. dr. ir. F. Mulder  
ten overstaan van een door het  
College voor promoties ingestelde commissie  
in het openbaar te verdedigen

op vrijdag 13 juni 2008 te Heerlen  
om 15.30 uur precies

door

Johan Anton van Buuren

geboren op 29 augustus 1946 te Culemborg

---

*Promotor*

Prof. dr. J. von Grumbkow, Open Universiteit Nederland

*Toegevoegd promotor*

Dr. M. van der Klink, Open Universiteit Nederland

*Overige leden beoordelingscommissie*

Prof. dr. M.P.F. Berger, Universiteit Maastricht

Prof. dr. G. Schuyten, Universiteit Gent

Prof. dr. H.T. van der Molen, Erasmus Universiteit Rotterdam, Open Universiteit Nederland

Prof. dr. G. Zwaneveld, Open Universiteit Nederland

## Inhoud

### Voorwoord

<b>1</b>	<b>Van problemen in het statistiekonderwijs naar het doel van de interventie</b>	<b>11</b>
1.1	Inleiding	11
1.2	Problemen in het statistiekonderwijs	12
1.2.1	Problemen in het statistiekonderwijs: de student als verklaringsfactor	12
1.2.2	Problemen in het statistiekonderwijs: het onderwijs als verklaringsfactor	14
1.2.3	Problemen in het statistiekonderwijs: de plaats van het statistiekonderwijs in het psychologiecurriculum als verklaringsfactor	15
1.3	Interventiedoel, vraagstelling en overzicht van het proefschrift	16
<b>2</b>	<b>Van interventiedoel naar veranderdoelen</b>	<b>21</b>
2.1	Inleiding	21
2.2	Bepaling van de gewenste leeruitkomsten: gedragsdoelen	22
2.3	Selectie van de determinanten	23
2.3.1	Direct veranderbare externe determinanten	25
2.3.2	Indirect veranderbare persoonlijke determinanten: procesvariabelen	32
2.4	Matrix van veranderingsdoelen	36
<b>3</b>	<b>Van veranderdoelen naar instructieaanwijzingen</b>	<b>39</b>
3.1	Inleiding	39
3.2	Theoretische methodieken, praktijken en instructieaanwijzingen	39
3.2.1	Cognitive Load Theory	39
3.2.2	Meaningful Learning	41
3.2.3	Cognitive Apprenticeship Instruction Model (CAIM)	42
3.2.4	Enkele empirische bevindingen van CAIM	44
3.3	Theory of the Research Training Environment	46
3.4	Matrix van instructieaanwijzingen	49
<b>4</b>	<b>Van instructieaanwijzingen naar een geïntegreerd competentiegericht onderwijsontwerp</b>	<b>51</b>
4.1	Inleiding	51
4.2	Het onderwijsontwerp	51
4.2.1	Four-Component Instructional Design Model (4C/ID)	52
4.2.2	Redenen voor de keuze van het ontwerpmodel	54

4.3	Naar een blauwdruk voor het M&T- en statistiekonderwijs	57
4.3.1	Onderzoekstaak als leertaak	57
4.3.2	Omvang en inhoud van benodigde M&T- en statistiekkennis	58
4.3.3	Onderwijsblauwdruk van het onderwijsprogramma voor de onderzoekscompetentie	58
4.3.4	Inpassing van het ontwerp van de onderzoekscompetentie in het curriculum van de opleiding psychologie	59
4.4	Pilots met het competentiegericht onderwijsontwerp	61
4.4.1	Inleiding	61
4.4.2	Inhoud en doelstellingen van de pilots	62
4.4.3	Methoden	65
4.4.4	Resultaten	66
4.4.5	Discussie en conclusie	73
<b>5</b>	<b>Van ontwerp naar implementatie: crossectionele bevindingen</b>	<b>77</b>
5.1	Inleiding	77
5.1.1	Onderzoeksvraag en hypothesen	77
5.2.	Methoden	78
5.2.1	Respondenten	78
5.2.2	Variabelen en hun operationalisering	79
5.2.3	Procedure	89
5.2.4	Analyseschema	90
5.3	Resultaten	91
5.3.1	Vergelijking van de twee leeromgevingen	91
5.3.2	Toetsing van de mediatie in het model met SEM	93
5.4	Conclusie en discussie	97
5.4.1	Conclusie effecten van competentiegericht versus vakgericht statistiekonderwijs	97
5.4.2	Beperkingen van het onderzoek	100
5.4.3	Slotbeschouwing	102
<b>6</b>	<b>Van ontwerp naar implementatie: longitudinale bevindingen</b>	<b>103</b>
6.1	Inleiding	103
6.1.1	Onderzoeksvraag	103
6.2	Methoden	104
6.2.1	Respondenten	104
6.2.2	Meetinstrumenten	104
6.2.3	Procedure	106
6.2.4	Analyseschema	107
6.3	Resultaten	107
6.4	Conclusie en discussie	110
6.4.1	Conclusie effecten van competentiegericht versus vakgericht statistiekonderwijs	110
6.4.2	Beperkingen van het onderzoek	112
6.4.3	Slotbeschouwing	113

<b>7</b>	<b>Evaluatie: reflectie en discussie</b>	<b>115</b>
7.1	Inleiding	115
7.2	Samenvatting en beantwoording van de onderzoeksvragen	115
7.3	Opbrengsten van het onderzoek	120
7.4	Beperkingen van het onderzoek	122
7.5	Aanbevelingen voor vervolgonderzoek	125
7.6	Aanbevelingen voor de praktijk van statistiekonderwijs nieuwe stijl	127
7.7	Tot slot	127
	 Bijlage I.0 Stappen Intervention Mapping	 129
	Referenties	133
	Samenvatting	147
	Summary	151
	Dankwoord	155
	Lijst van eigen publicaties en presentaties in relatie tot het thema van dit proefschrift	157

#### *Cd-rom*

*Van vakgericht naar competentiegericht statistiekonderwijs*

#### Bijlagen

##### **I Matrices**

- I.0 Stappen Intervention Mapping
- I.1 Matrix van veranderdoelen
- I.2 Matrix van instructieaanwijzingen
- I.3 Matrix tien ontwikkelstappen van het 4C/ID-model met instructieaanwijzing en
- I.4 Onderwijsblauwdruk onderzoekscompetentie bachelorfase OUNL (situatie 2007)

##### **II Meetinstrumenten Vragenlijst**

- II.1 Vragenlijst evaluatie IMTO-pilots
- II.2 Vragenlijst evaluatie onderzoekscompetentie 2005
- II.3 Vragenlijst evaluatie T<sub>1</sub>-meting

##### **III Voorbereidende analyses**

- III.1 Vervolgonderzoek naar nonrespons in het crosssectionele onderzoek
- III.2 Voorstudie Raschanalyses van de oorspronkelijke Likertschalen
- III.3 Kruisvalidatie van het determinantenmodel met Structural Equation Modeling
- III.4 IJking T<sub>1</sub>-T<sub>2</sub> -items met Raschmodel





## Voorwoord

Tien jaar geleden vond bij de opleiding Psychologie van de Open Universiteit Nederland de omslag plaats van vakgericht naar competentiegericht statistiekonderwijs. Tot dat moment werd er door mij voortdurend gewerkt aan de optimalisering van het statistiekonderwijs voor psychologiestudenten waarbij de inhoud van het vak steeds centraal stond. De strategie die ik samen met Bert Nijdam ontwikkelde, reikte op dat moment nog niet verder dan het ordenen van de statistiekinhouden aan de hand van wetenschappelijke onderzoeksvragen.

Bij de Open Universiteit Nederland hebben Bert Nijdam en ik de statistiekt teksten vervolgens toegesneden op het afstandsonderwijs. De komst van ICT vergrootte het repertoire aan onderwijskundig handelen in het toegankelijk maken van de statistiekinhouden (tweedegeneratie-onderwijs). Een voorbeeld daarvan is de Interactieve Leer- en Cursusontwikkelomgeving (ILCO). ILCO streefde twee doelen na. Ten eerste moest dit programma tegemoetkomen aan de vraag naar adaptatie van leermaterialen door de student. Meer dan een reguliere universiteit kent de Open Universiteit Nederland een grote verscheidenheid van studenten die alle met dezelfde cursus uit de voeten moeten kunnen. Door het kennisbestand te splitsen in basisstof (domeinspecifieke inhoud) en begeleidingscomponenten (zoals voorbeelden, illustraties, samenvattingen, schema's, advanced organizers, opdrachten en vraagstukken) die gemiddeld 45% van het leermateriaal uitmaken, kunnen de OUNL-studenten nu zelf invloed uitoefenen op de leermaterialen (in termen van hoeveelheid basisstof en ingebouwde begeleiding). Een vak als statistiek leent zich uitstekend voor zo'n opdeling en uitwerking. Midden jaren negentig is in een aantal pilots deze methode beproefd. Een tweede doel kwam tegemoet aan de vraag van de onderwijsorganisatie naar een grotere efficiëntie in het proces van cursusontwerp, -productie en -exploitatie. Door de opkomst van het competentiegericht denken is aan een beoogd opvolger van ILCO op het internet – Mercator – niet meer toegekomen.

Eveneens verliep de presentatie van het onderwijs in de methoden en technieken van onderzoek (M&T), dus tot voor tien jaar, nog langs de lijnen van vakspecifieke inhouden. Ook daar heb ik naast het schrijven van teksten enkele elektronische hulpmiddelen ontwikkeld om studenten te instrueren hoe onderzoek aangepakt moet worden. Een voorbeeld daarvan is het programma Onderzoeksmanager. Daarnaast introduceerde ik in 1997 Onderzoeker de basis. Met dit op cd-rom uitgeleverde programma worden studenten via een checklistmethode door allerlei onderzoeksopzetten met bijbehorende dataverzamelings- en analysetechnieken genavigeerd.

Eind jaren negentig krijgt het competentiegericht denken vaste voet aan de grond bij de Open Universiteit Nederland. Dit valt samen met de versnelde verspreiding van ICT. Beide ontwikkelingen maakten het mogelijk de vakken statistiek en M&T niet meer los van elkaar en enkel vanuit de desbetreffende vakstructuren te behandelen.

De studenten moeten zich nu al onderzoekende de inhouden van deze vakken eigen maken. Elke academicus mag immers aangesproken kunnen worden op zijn of haar vermogen – competentie – om kennis te genereren of toe te passen. Dit houdt niet in dat elke academicus een onderzoeker moet zijn, maar wel dat deze als ‘scientist practitioner’ zijn of haar professie beoefent. In plaats van losse cursussen statistiek en M&T worden studenten nu vanaf het begin van de studie uitgenodigd, met een cumulerende reeks van steeds complexer wordende onderzoeken, zich als onderzoeker de kennis en vaardigheden uit het domein van de psychologie, M&T én de statistiek eigen te maken. De verwachting is dat de student zo op een natuurlijke wijze vertrouwd raakt met het toepassen van statistiek in een verscheidenheid van psychologische contexten, de bijbehorende juiste methoden leert gebruiken en de verschillende bijpassende technieken weet te selecteren. Door statistiek te verbinden met de psychologie krijgt statistiek meer betekenis voor de psychologiestudent. En omgekeerd: de inhoudelijke vakken krijgen meer betekenis door hun integratie met statistiek. Voor meer informatie over de ontwikkelingen van het competentiegericht onderwijs bij de opleiding Psychologie van de Open Universiteit Nederland verwijst ik graag naar de website <http://www.cope.ou.nl/>. Dit proefschrift beschrijft de ontwikkeling van het competentiegericht ontwerp voor het statistiekonderwijs en het testen van de effectiviteit van het eerste geactualiseerde deel ervan.

Heerlen, 22 april 2008

Hans van Buuren

## **Van problemen in het statistiekonderwijs naar het doel van de interventie**

### **1.1 Inleiding**

Alle universitaire psychologieopleidingen tellen een of meer cursussen (toegepaste) statistiek in hun curricula. En alle kampen ze met dezelfde problemen. Studenten klagen dat ze niet begrijpen wat statistiek met psychologie te maken heeft en vinden het, los daarvan, moeilijk en saai. Statistiekdocenten doen er van alles aan om het vak aantrekkelijk en beter studeerbaar te maken, desondanks doen studenten geen moeite om 'hun' vak te willen beheersen. Psychologiedocenten klagen dat studenten bij hun doctoraalscriptie of masterthesis niet weten hoe ze hun data moeten analyseren. Dat hadden ze toch geleerd bij statistiek? Beleidsmedewerkers en het management klagen dat de slaagpercentages te laag zijn. Hierdoor stremt de doorstroming en zijn de rendementen te laag. Kortom: het lijkt erop dat niemand tevreden is met de huidige statistiekcursussen. Is er voor deze problematiek een oplossing die een einde maakt aan de blijvende onvrede? In dit proefschrift wordt gezocht naar een oplossing die buiten de kaders van het vak statistiek zelf valt. Als men zich realiseert dat statistiek een functie heeft binnen het grotere geheel van 'het doen van (psychologisch) onderzoek', waarom dan niet het doen van (psychologisch) onderzoek als meeromvattende 'hele taak' als uitgangspunt voor het onderwijs nemen? In theorie is dit een eenvoudige oplossing, maar in de praktijk niet of nauwelijks te verwezenlijken. Dus handhaven psychologieopleidingen de benadering om studenten in het doen van psychologisch onderzoek te scholen door afzonderlijke ingrediënten van die bekwaamheid apart aan te bieden, in het bijzonder de vakken methoden en technieken van onderzoek (M&T) en statistiek. De studenten moeten zelf maar zien hoe zij die vakken integreren en toe gaan passen in relatie tot het beantwoorden van een psychologische onderzoeksvraag voor hun bachelor- of masterthesis. Binnen de opleiding Psychologie van de Open Universiteit Nederland echter is ervoor gekozen het vak statistiek volledig te laten opgaan in het nieuwe 'vak' de onderzoekscompetentie.

In dit hoofdstuk wordt eerst nader ingegaan op de problemen binnen het statistiekonderwijs (paragraaf 1.2), eerst vanuit het studentperspectief (paragraaf 1.2.1), dan vanuit het onderwijsperspectief (paragraaf 1.2.2) en ten slotte vanuit het curriculumperspectief (paragraaf 1.2.3). In paragraaf 1.2.4 wordt aangegeven wat de speelruimte is om veranderingen in het statistiekonderwijs en in het curriculum aan te brengen. Vervolgens belicht paragraaf 1.3 hoe de Open Universiteit Nederland de problemen met statistiek in de opleiding Psychologie denkt op te lossen. Het resultaat is een interventie die leidt tot de herinrichting van een aantal vakken in het curriculum van de opleiding Psychologie. Voor de reconstructie en beschrijving van het interventie-

proces wordt in dit proefschrift gebruik gemaakt van de systematiek van Intervention Mapping (Bartholomew, Parcel, Kok, & Gottlieb, 2006).

## 1.2 Problemen in het statistiekonderwijs

Welke zijn de problemen die zich in het statistiekonderwijs voordoen en waaraan worden die problemen in het veld toegeschreven? Hierna volgt een uiteenzetting van de problemen vanuit een student-, een onderwijs- en een curriculumperspectief.

### 1.2.1 *Problemen in het statistiekonderwijs: de student als verklaringsfactor*

De problemen in het statistiekonderwijs aan aankomende psychologen die enkele decennia terug werden genoemd spelen nog volop: 'Statisticophobie' (Beenen & Van der Werff, 1967; Dillon, 1982), 'ongecijferdheid' (Chalmers & Hill, 1973), gebrek aan motivatie (Kalton, 1973), moeite met het leren van statistiek (Garfield & Ahlgren, 1988), gebrek aan retentie (Clark, Karuat, Mathews, & Wimbish 2003; Garfield & Ben-Zvi, 2007; Mathews & Clark, 2003; Willems, 1981) en het onvermogen om data te analyseren bij eenvoudig empirisch onderzoek (Garfield & Ben-Zvi, 2007; Imbos & Dekeyser, 1999; Kneefel & Van der Leeden, 1999; Lagerwaard & Mul, 1982; Oost & De Jong, 1996; Salsburg, 1985; Wynne, Lockhorst & Westers, 1999). Sommige statistiekdocenten en onderzoekers van het statistiekonderwijs zijn geneigd de oorzaken daarvan bij de student te leggen. Daarbij hanteert men argumenten die verwijzen naar hun (deficiënte) vooropleiding, hun (negatieve) attitude ten aanzien van de statistiek en hun verwachtingen bij de aanvang van de studie. Hieronder worden deze argumenten kort toegelicht.

#### *Deficiënte vooropleiding*

Veel studenten die zich aanmelden voor een studie psychologie blijken onvoldoende wiskunde te hebben gehad om in het wetenschappelijk voortgezet onderwijs probleemloos statistiek te kunnen volgen. Daardoor vinden veel studenten statistiek een moeilijk vak, waarvoor zij harder moeten werken dan voor veel andere vakken. Sijsma en Van Duijn (1999), concluderen dat: '... statistiek zo ongeveer de status [lijkt] te hebben van wiskunde op de middelbare school: moeilijk en niet leuk, en bovendien valt niet zo gemakkelijk in te zien wat je er nu precies aan hebt' (p. 5). Statistiekdocenten zien zich geplaagd voor grote groepen studenten die sterk heterogeen in vooropleiding zijn (Moonen, 1978; Tempelaar, 2002, 2004). Sijsma en Van Duijn (1999) illustreren de heterogeniteit van de diverse vooropleidingen met de volgende situatie: na een college statistiek komt eerst een student langs die klaagt over het trage tempo en de 'middelbare' schoolstof; daarna komt een andere student die wel zijn waardering uit voor het college, maar eraan toevoegt dat het wel allemaal abracadabra blijft.

#### *Negatieve attitude en angst ten aanzien van de statistiek*

Docenten constateren dat veel studenten zich van meet af aan negatief opstellen ten aanzien van het vak statistiek en angst voor het vak hebben (Baloğlu, 2003; Baloğlu

& Zelhart, 2003; Beenen & Van der Werff, 1967; Berk, 1996; Gordon, 2004; Lalonde & Gardner, 1993; Schau, Stevens, Dauphinee, & Del Vecchio, 1995). De angst speelt zowel voor, tijdens als na het volgen van de statistiekcursussen (Gal & Ginsburg, 1994). Onwuegbuzie en Daley (1999, p. 1091) stellen: 'Statistics anxiety is an anxiety which occurs as a result of encountering statistics in any form and at any level (...) which appears to involve a complex array of emotional reactions which have the propensity to debilitate learning'. Hier wordt ook wel gesproken van 'statisticofobie' (Beenen & Van der Werff, 1967; Dillon, 1982) of 'sadistics' (Rosenthal, 1992). Deze angst wordt mogelijk ontwikkeld of versterkt doordat statistiek in het verleden niet zelden heeft gediend als een verkapt selectiemechanisme (Verschuren, 2002).

#### *Verwachtingen en waardering*

Studenten onderkennen in de eerste fase van hun studie, wanneer het statistiekonderwijs begint, nog niet het belang van dit vak voor het wetenschappelijk onderzoek dat zij in een latere fase van hun opleiding geacht worden uit te voeren (Gelso, 2006; Moonen, 1978; Wynne, Lockhorst, & Westers, 1999). Zij beseffen niet dat zij zich tijdens hun studie dienen te transformeren van consument van psychologische kennis naar producent daarvan (Donald, 2002). Voor veel psychologiestudenten is het vak statistiek een noodzakelijke en vooral lastige horde op weg naar de mastertitel (Perney & Ravid, 1991, in Gal & Ginsburg, 1994). Velen vragen zich af wat de relevantie is van een vak als statistiek voor bijvoorbeeld de psychologie of de geneeskunde. 'Mensen zijn toch geen cijfers of getallen?'.

Ongemotiveerde en angstige studenten met een negatieve attitude passen oppervlakkige leerstrategieën toe die niet lijken bij te dragen tot een beter begrip van de statistiek: uit het hoofd leren en vlak voor het examen gaan pieken (Wynne, Lockhorst, & Westers, 1999). De consequentie is dat de stof niet beklijft en dus niet toegepast kan worden (Gelso, 2006; Lagerwaard & Mul, 1982; Oost & De Jong, 1996; Salsburg, 1985; Schmidt, 1997; Verschuren, 2003; Willems, 1981).

Doorgaans blijkt dat studenten bij de analyse van onderzoeksdata in een later stadium van hun studie (bachelorthesis of masterthesis) aangeven dat ze eigenlijk niet weten hoe ze dit moeten aanpakken (Salsburg, 1985). Blijkbaar treedt er onvoldoende 'transfer of training' op (Tomic & Span, 1993). Onderzoek wijst erop dat zelfs de 'gemiddelde' student in het geheel niet statistisch kan denken over belangrijke psychologische onderwerpen (Shaughnessy, 1992; Kuhn, 1991; Konold, 1989; Tversky & Kahneman, 1974), dat statistische begrippen vaak amper of zelfs verkeerd begrepen worden (Garfield & Ahlgren, 1988; Shaughnessy, 1992) en als dan die 'gemiddelde student' probeert statistische methoden te gebruiken '...the results are a shambles' (Butler, 1998, p. 84). Men kan zich afvragen wat studenten eigenlijk leren tijdens de statistiekcursussen. Onderzoek van Mathews en Clark (2003) en Clark, Karuat, Mathews en Wimbish (2003) bij hoogpresterende studenten (met een Grade A voor hun eerste statistiekcursus) in vier verschillende universiteiten heeft uitgewezen dat zelfs deze studenten binnen zes weken na de succesvolle afsluiting van de cursus een gebrekkig begrip hebben van het gemiddelde en de standaarddeviatie.

Een gevolg van de oppervlakkige leerstrategieën is dat studenten niet leren empirische uitspraken kritisch te benaderen waardoor ook hun academische vorming als ‘scientist practitioner’ tekortschiet. Hiermee wordt bedoeld dat de beroepsuitoefening van de praktiserende psycholoog ‘evidence based’ dient te zijn, dat wil zeggen: gebaseerd op de actuele stand van zaken in het empirische onderzoeksveld en op zorgvuldige empirische evaluatie van het eigen beroepsmatige handelen (Commissie Levelt, 2006; Gelso, 2006).

### 1.2.2 *Problemen in het statistiekonderwijs: het onderwijs als verklaringsfactor*

Het voorafgaande stelt dat de oorzaken voor de problemen in het statistiekonderwijs voor een deel terug te voeren zijn op kenmerken van studenten. Daarnaast speelt ook het onderwijs zelf een rol van betekenis. Hieronder wordt ingegaan op problemen inzake de inhoud en de didactiek van het statistiekonderwijs.

#### *Inhoud*

Hoewel statistiek niet gezien moet worden als een wiskundevak (Moore, 1997, zie ook hierna) hebben ze wel enkele punten gemeen. Ten eerste kent het vak in de opbouw van zijn begrippenapparaat net als de wiskunde een duidelijk hiërarchische structuur. Deze hiërarchie vergroot de kansen op cumulatief opgebouwde problemen bij studenten voor wie de behandeling van die begrippen reeds vanaf het begin van de studie te snel gaat. Statistische kernbegrippen als ‘variabele’ en ‘variatie’ maar ook ‘verband’ blijken studenten tot aan de fase van de masterthesis voor problemen te stellen. Een deel van de ervaren problemen met statistische begrippen heeft te maken met de in wiskundige termen gedefinieerde en afgeleide begrippen (vergelijk het gebruik van definitieformules en rekenformules voor allerlei statistische maten en technieken). Onvoldoende wiskundige voorkennis bij een grote groep studenten belemmert dan een ‘snel’ begrip van die maten en technieken. Vooral de kansrekening, waarin het beroep op de wiskundige voorkennis tamelijk groot is, stelt veel studenten voor problemen. Als er dan ook nog een veelheid van statistische toetsen volgt, overzien studenten vaak het geheel niet meer.

#### *Didactiek*

Docenten zijn in hun onderwijs gebonden aan de statistische inhoud en de structuur van deze discipline. Deze moeten in een beperkte tijd worden overgedragen. Dit komt erop neer dat studenten eerst veel begrippen en de relaties daartussen moeten leren alvorens ze te kunnen toepassen. De huidige ontwikkelingen in ICT bieden de docenten de mogelijkheid om bepaalde handelingen zoals het tijdrovende rekenwerk achterwege te laten en de vrijgekomen ruimte met meer toepassingen in te vullen. In de USA heeft dit geleid tot een beweging onder statistiekdocenten waarvan de boodschap luidt: ‘More data and concepts, less theory, fewer recipes’ (zie onder anderen Cobb, 1992; Moore, 1997). In de lijn met recente didactische principes van het cognitivisme en het constructivisme krijgen statistiekdocenten het dringende verzoek meer activerende opdrachten in hun lessen aan te bieden en minder uit te gaan van passieve



(hoor)colleges (Gal & Garfield, 1997). Langzamerhand verschuift de aandacht in de behandeling van de statistische methoden van definities, afleidingen, bewijsvoering en rekenvaardigheid in de richting van het waarom, het wanneer en het hoe van het gebruik van deze methoden (Moore, 1997). In de onderzoeksliteratuur komt men daarvoor de term ‘statistisch denken’ tegen (Britz et al., 1997; Ben-Zvi & Garfield, 2004; Chance, 2002; Moore, 1997; Wild & Pfannkuch, 1999; Ullman, 1996). Ondanks deze aanbevelingen blijken er in de praktijk nog relatief weinig veranderingen doorgevoerd te zijn. Garfield (1997) constateert dat in veel onderwijsinstellingen studenten nog steeds passief in grote collegezalen hun docenten moeten aanhoren. Het enige verschil is dat luchtige powerpointpresentaties schoolbord en krijt verdrongen hebben: ‘Statistics teaching is often stagnant; statistics teachers resist change’ aldus Hogg (1992, p. 36).

Volgens Moore (1997) vindt het meest effectieve leren plaats wanneer *inhoud* (datgene wat we studenten willen laten leren), *didactiek* (wat we doen om hen te helpen leren) en *technologie* (welke middelen we daarbij gebruiken) elkaar wederzijds versterken. Specialisten op elk der drie terreinen neigen ertoe het belang van de andere terreinen te onderschatten (Moore, 1997). Moore werkt zijn opvatting, die vraagt om een belangrijke organisatorische ingreep, echter niet uit in concrete lesprogramma’s. Het gebrek aan synergie tussen inhoud, didactiek en technologie (Moore, 1997) blijft voortbestaan. Cases worden nu vaker aangeboden en besproken in kleinere werkgroepen en statistiek wordt geïncorporeerd in leeronderzoeken. Dit leidt weliswaar tot wat meer waardering van studenten voor het vak statistiek (Attema, 1996), maar haalt de problemen bij grote groepen studenten niet weg (Engels-Freeke & Flach, 1999; Garfield, 1997; Imbos & Dekeyser, 1999; Sijtsma & Van Duijn, 1999; Wynne, Lockhorst, & Westers, 1999). Ongemotiveerdheid, angst, een negatieve houding en een oppervlakkige studeeraanpak blijven hardnekkig voortbestaan. Dat betekent ook dat de docenten vaak met een gefrustreerd gevoel achterblijven. Het draagt niet bij aan arbeidsvreugde van de statistiekdocent om als docent geconfronteerd te worden met gebrek aan motivatie, tekortkomingen in wiskundige vaardigheden, onregelmatig college- en werkgroepbezoek, slechte prestaties en lage slaagpercentages bij tentamens en last but not least klachten van collega’s die de statistiekdeficiënties bij studenten in hun bachelor- en masterthesisfase opmerken.

### 1.2.3 *Problemen in het statistiekonderwijs: de plaats van het statistiekonderwijs in het psychologiecurriculum als verklaringsfactor*

In alle psychologieopleidingen wordt het vak statistiek gekenschetst als een hulpdiscipline, meestal verzorgd door statistici. Niet-statistici zoals psychologen met een specialisatie in het statistiekonderwijs conformeren zich in het algemeen aan de normen die de statistici stellen (Imbos & Dekeyser, 1999). De organisatorische positionering van de statistiek binnen opleidingen als die van de psychologie leidt ertoe dat het vak statistiek binnen die opleidingen een eigen leven leidt.

De inhouden van de statistiekcursussen zijn geordend op basis van statistische methoden. Volgens Bakker (2004) worden die inhouden los van elkaar behandeld:



‘topic-topic-topic approach in which topics are addressed without coherence’ (p. 276). Meestal wordt statistiek geïntroduceerd als (a) een overzichtscursus, (b) georganiseerd volgens statistische methoden en begrippen, (c) gepresenteerd in een standaardvolgorde, (d) met de docent die bijna steeds het woord voert (Cobb, 1993, p. 9 en verder). Geen van deze vier kenmerken is wezenlijk voor inleidende cursussen statistiek in het wetenschappelijk onderwijs. Data en cases uit het domein van de psychologie dienen eerder als illustratie van het gebruik van statistische methoden dan als uitgangspunt om data te analyseren met als doel het beantwoorden van psychologisch relevante onderzoeksvragen (Cobb, 1993).

Het gevolg van de zelfstandige en daarmee geïsoleerde plaats van het statistiekonderwijs is dat statistiek door studenten als ‘losstaand’ vak gezien wordt, dat in de praktijk geen of nauwelijks raakvlakken lijkt te hebben met de overige vakken (Imbos & Dekeyser, 1999, p. 10). De beleving dat de statistiek los lijkt te staan van de rest van de opleiding leidt ertoe dat studenten moeite hebben met het integreren van deze kennis en vaardigheden in het algemene kennis- en vaardighedenbestand dat vereist is om hun beroep uit te oefenen. Losstaande feitenkennis wordt bovendien snel vergeten (Dolmans & Schmidt, 1996; Roorda & Van Streun, 2002). Dit kan deels verklaren waarom de studenten het nut van de statistiek niet zien. Verder stellen onder anderen Verschuren (2002) en (Gelso (2006) dat studenten een verkeerd beeld van onderzoek krijgen indien ze met statistiek als eerste methodologievak geconfronteerd worden. Verschuren noemt de volgende aspecten die volgens hem veel meer de einduitslag van een onderzoek bepalen dan de data-analyse: de analyse van het probleem, de vaststelling van de kennis die nodig of nuttig is bij de oplossing en de vertaling hiervan in een serie adequaat sturende onderzoeksvragen, de selectie van de theorieën, de definiëring van de centrale concepten en de vertaling ervan in waarnemingstermen en ten slotte de gekozen onderzoeksstrategie (Verschuren, 2002).

### 1.3 **Interventiedoel, vraagstelling en overzicht van het proefschrift**

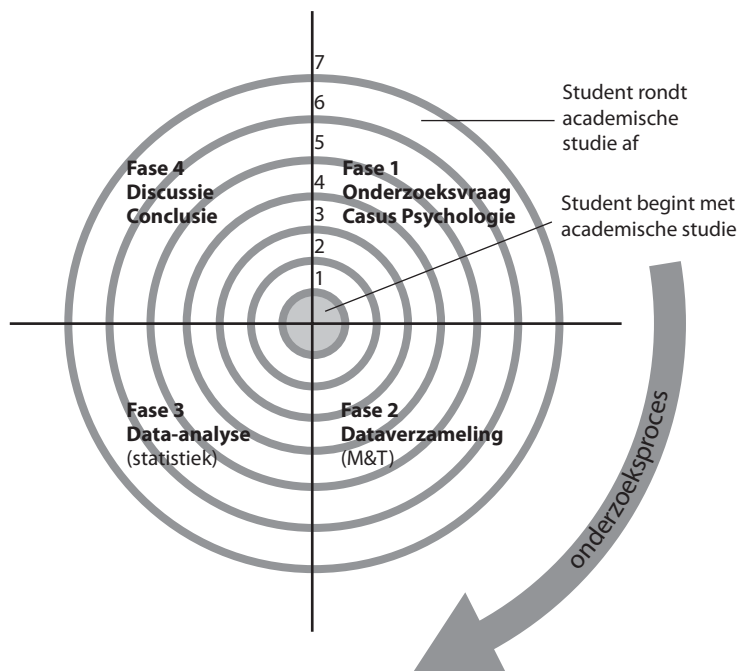
In het voorgaande zijn vanuit het perspectief van de student, het onderwijs (de docent) en het curriculum de problemen met het statistiekonderwijs geschetst. Om een oplossing voor die problemen te vinden, moet duidelijk worden welke factoren van dusdanig belang én te beïnvloeden zijn dat daarmee de problemen verdwijnen dan wel verminderen. In dit proefschrift wordt ingegaan op de derde verklaring die is gegeven, te weten: de plaats die het statistiekonderwijs in het psychologiecurriculum inneemt. De stelling wordt gehuldigd dat de problemen niet binnen de grenzen van het vakgebied zelf zijn op te lossen. Het is de positionering van het statistiekonderwijs in het psychologiecurriculum die wordt beschouwd als de primaire veroorzaker van de problemen in het statistiekonderwijs. Door de positionering van het vak statistiek in het curriculum te veranderen, kan worden bereikt dat aanpassingen in statistiekinhouden en de wijze van overdracht (didactiek) mogelijk zijn die meer zijn toegesneden op de behoeften en capaciteiten van de student. De verwachting is dat daardoor beïnvloedbare studentkenmerken zoals attitude, angst, motivatie, waarde-

ring en studeergedrag die in het veld worden gezien als medeveroorzakers van de problemen, in positieve zin ombuigen.

Bij de opleiding Psychologie van de Open Universiteit Nederland heeft de faculteit besloten studenten vanaf het begin van de studie te confronteren met een onderzoeksbenadering. Studenten kunnen zich in een cumulerende reeks van steeds complexer wordende onderzoekingen al onderzoekende de kennis en vaardigheden uit het domein van de psychologie, methoden en technieken van onderzoek (M&T) én de statistiek eigen maken. De verwachting is dat de student op deze manier op een natuurlijke wijze vertrouwd raakt met het toepassen van statistiek in een verscheidenheid aan psychologische contexten, de bijbehorende juiste methoden leert gebruiken en de verschillende bijpassende technieken weet te selecteren.

Door statistiekcursussen te integreren met delen uit psychologiecursussen en M&T-cursussen wordt de mogelijkheid benut de statistiekleerstof over een groter aantal cursussen te verdelen dan de gebruikelijke één à twee, waardoor de complexiteit van de leerstof en dus de cognitieve belasting per cursus vermindert. Door statistiek te verbinden met de psychologie krijgt statistiek meer betekenis voor de psychologiestudent. En omgekeerd: de inhoudelijke vakken krijgen meer betekenis door hun integratie met statistiek. De onderzoekstaken beogen een bijdrage te leveren aan een verbeterde motivatie voor het doen van onderzoek en een positievere attitude ten aanzien van statistiek. Meer ervaring met de onderzoekspraktijk draagt tevens bij aan het academische ideaalbeeld van de praktiserend psycholoog als ‘scientist practitioner’ (Commissie Levelt, 2006; *Rapport Onderwijsvisitatie Psychologie*, 2007). Met een dergelijk competentiegericht ontwerp wordt statistiek een onlosmakelijk deel van de onderzoekscompetentie. Met het herontwerpen van een gedeelte van het psychologiecurriculum door de invoering van de onderzoekscompetentie als oplossing van de problemen in het statistiekonderwijs is het doel van de interventie omschreven.

Figuur 1.1 symboliseert de opbouw van de onderzoekscompetentie met de M&T- en statistiekleerstof in een aantal te doorlopen onderzoeksronden. Het onderzoeksproces is onderverdeeld in vier fasen die overeen komen met de indeling in secties of fasen die door de American Psychological Association (APA, *Publication Manual*, 2002) voor het format van artikelen in wetenschappelijke tijdschriften wordt voorgeschreven: ‘Introduction’, ‘Methods’, ‘Results’ en ‘Discussion’. In deze fasen zijn ook de afzonderlijke vakken te herkennen. De buitenste ring geeft het niveau van de onderzoekscompetentie aan waarop studenten geacht worden voldoende kennis en vaardigheden te hebben opgedaan om hun bachelorthesis tot een goed einde te brengen (niveau 7). Vanuit de eisen die aan de thesis worden gesteld, wordt teruggeredeneerd welke kennis en vaardigheden studenten op het terrein van de statistiek, M&T en psychologie zich dienen eigen te maken. De inhoud van deze vakken kunnen dan op voor studenten meer betekenisvolle wijze over de verschillende ringen worden verdeeld. Bij de opleiding Psychologie moeten de inhoud van twee statistiekcursussen, een cursus SPSS en drie cursussen M&T worden herverkaveld over de zes binnenringen in de vorm van onderzoeksrondes.



FIGUUR 1.1 Onderzoekscompetentiecirkel

In dit proefschrift wordt verslag gedaan van de wijze waarop de onderzoekscompetentie gestalte heeft gekregen en welke resultaten dit heeft opgeleverd. Bij het ontwerpen zijn diverse disciplines betrokken. Te denken valt aan psychologische principes over het leren van abstracte en complexe kennis, het opwekken van motivatie, het veranderen van attitudes, onderwijskundige richtlijnen en ontwerpen van onderwijs die alle bijdragen aan gewenste leeruitkomsten. De onderzoeksvraag luidt:

*In hoeverre leidt competentiegericht onderwijs (de interventie) tot een effectieve leeromgeving waarin de afstand tussen gewenste en gerealiseerde leerdoelen ten aanzien van begrip en toepassing van de statistiek in psychologisch onderzoek kleiner is dan in het vakgerichte onderwijs?*

In 1998 is een begin gemaakt met de voorbereidingen van de interventie (Van Buuren & Giesbertz, 1998). De implementatie van de interventie zal in 2008 grotendeels zijn afgerond. Voor een *reconstructie* van de interventie wordt in dit proefschrift gebruik gemaakt van Intervention Mapping (Bartholomew, Parcel, Kok, & Gottlieb, 2006). Deze interventiebenadering is afkomstig uit de wereld van de planmatige gezondheidsvoorlichting. De benadering ontleent haar attractiviteit aan haar instrumentarium, de zogenoemde transitie matrices, die een beknopte en transparante beschrijving van het interventieontwikkelingsproces mogelijk maken. Bovendien verduidelijken de matrices hoe het ontwerp op systematische wijze kan worden verankerd in em-

pirische gegevens en theoretische kennis over het veranderen van gedrag (vergelijk de kwalitatieve matrices bij Miles & Huberman (1984; 1994) en bij Hutjes en Van Buuren (1992)). Door het gebruik van de matrices resulteert Intervention Mapping uiteindelijk in een plattegrond of routeplanner voor het te ontwerpen programma i.c. de herinrichting van een gedeelte van het psychologiecurriculum. In bijlage I.0 van dit proefschrift zijn de opeenvolgende stappen – van de behoeftebepaling tot de evaluatie van de implementatie van een ontwerp van de interventie – met in elkaar overlopende matrices en tabellen uitgewerkt. Elke stap wordt in dit proefschrift door een hoofdstuk vertegenwoordigd. Met de beschrijving van de problemen in het statistiekonderwijs in het voorgaande is de eerste taak in interventiestap 1, de behoeftebepaling, uitgevoerd. Wellicht ten overvloede: het ontwerpproces zelf is in dit proefschrift niet volgens Intervention Mapping uitgevoerd. De methode dient hier de systematische reconstructie van het ontwerpproces en de reflectie daarop, die moet leiden tot het traceren van de hiaten en geleerde lessen vanuit de ontwerppraktijk

Het doel van de interventie kent de volgende beperkende randvoorwaarden. Een eerste beperking heeft te maken met het type onderzoek waarin de statistiek is ingebed. In dit proefschrift wordt gefocust op het onderzoeksmodel volgens de empirische cyclus van De Groot (1961), dus het genereren en toetsen van (psychologische) kennis. Verder beperkt het onderwijsontwerp zich tot de bachelorfase van de opleiding Psychologie. Een derde beperking betreft het type onderwijsinstelling: het onderzoek vindt plaats aan de Open Universiteit Nederland, een instelling voor afstandsonderwijs. Dit type onderwijs kent géén campus van studenten en beperkte directe docent-student (face-to-face) contacten. Het cursusmateriaal moet op zich voldoende zijn om het studeren onafhankelijk van tijd, plaats en tempo mogelijk te maken. De rol van de docent is als het ware ingebed in het cursusmateriaal zelf ('ingebouwde begeleidingscomponenten', Martens, 1998). Een vierde beperking heeft te maken met de *evaluatie* van de interventie: hoewel de onderzoekscompetentie ook aspecten van psychologische theorie en methoden en technieken van onderzoek inhoudt, wordt in dit onderzoek de nadruk gelegd op de statistische aspecten die daarin een rol spelen.

De verdediging van de basisstelling dat het competentiegericht statistiekonderwijs effectiever is dan het vakgericht statistiekonderwijs, is opgebouwd in de vorm van vijf deelvragen. Deze deelvragen lopen parallel aan de opeenvolgende stappen in het interventieproces volgens de IM-benadering.

- 1 Welke determinanten bepalen het succes van het leren van statistiek?
- 2 Welke onderwijskundige instructieaanwijzingen zijn effectief om de gewenste veranderingen in de leeromgeving te bewerkstelligen die leiden tot succes van het leren van statistiek?
- 3 Hoe kunnen de instructieaanwijzingen worden gesynthetiseerd tot een competentiegericht ontwerp dat een effectieve leeromgeving voor de onderzoekscompetentie inhoudt en hoe kan dat competentiegericht ontwerp worden geïmplementeerd in een bestaand, vakgericht curriculum?

- 4 *In welke mate is de competentiegerichte leeromgeving effectiever in het realiseren van leerprocessen en leeruitkomsten dan de vakgerichte leeromgeving?*
- 5 *In welke mate is de competentiegerichte leeromgeving op langere termijn effectiever qua effect op motivatie, leerstrategieën en statistiekkennis dan de vakgerichte leeromgeving?*

Hoofdstuk 2 gaat nader in op deelvraag 1. Na de specificatie van leerdoelen in concrete gedragingen (in Intervention Mapping termen: ‘performance objectives’ of gedragsdoelen) wordt nagegaan door welke determinanten de gewenste gedragingen te beïnvloeden zijn. Gedrags- en omgevingsfactoren die de problemen in het statistiekonderwijs veroorzaken worden geherformuleerd in gedrags- en omgevingsfactoren die het statistiekonderwijs effectiever maken. Gedragsdoelen en hun determinanten leiden tot een overzicht van veranderdoelen (in Intervention Mapping termen: ‘change objectives’) die de basis vormen voor de interventie in de vorm van een gedeeltelijke herinrichting van het psychologiecurriculum. In bijlage I.1 zijn gedragsdoelen, determinanten en veranderdoelen samengevat in een matrix.

Hoofdstuk 3 geeft een overzicht van de instructieaanwijzingen (in Intervention Mapping termen: ‘theory-based methods’ en ‘practical strategies’) die aansluiten op de veranderdoelen en die een bijdrage leveren aan een effectieve leeromgeving. De instructieaanwijzingen zijn gebaseerd op empirisch evidente onderwijskundige theorieën en praktijken. In bijlage I.2 is de matrix van instructieaanwijzingen te vinden. Hoofdstuk 4 beschrijft de wijze waarop de instructieaanwijzingen in een coherent competentiegericht onderwijsontwerp zijn samengevoegd en in een bestaand – vakgericht – curriculum is geïntegreerd. Het product in de vorm van een onderwijsblauwdruk voor de onderzoekscompetentie is te vinden in bijlage I.4. In een aantal pilots is de werkbaarheid van dit onderwijsontwerp verkend. De neerslag daarvan is eveneens in dit hoofdstuk opgenomen.

De hoofdstukken 5 en 6 beschrijven de resultaten van de implementatie van het nieuwe onderwijsontwerp in het bestaande curriculum van de opleiding Psychologie van de Open Universiteit Nederland. De resultaten van de nieuwe en de oude situatie bij de doelgroep van studenten worden met elkaar vergeleken. Hoofdstuk 5 bevat een crosssectioneel survey onder studenten die hetzij statistiek in het vakgerichte hetzij in het competentiegerichte onderwijs hebben gevolgd. Omdat ten tijde van dit onderzoek het nieuwe ontwerp nog niet integraal is ingevoerd, heeft deze evaluatie betrekking op de eerste onderdelen van het onderwijsontwerp waar de statistiek de hoofdmoot vormt. Deze wordt in figuur 1.1 gerepresenteerd door de binnenste ring (niveau 1) (deelvraag 4). In hoofdstuk 6 wordt een longitudinaal survey besproken. Ook hier worden twee groepen studenten vergeleken. Ook deze heeft betrekking op de statistiekkennis op niveau 1 (deelvraag 5).

Het laatste hoofdstuk geeft een samenvatting van de voornaamste bevindingen, een discussie en voorstellen voor vervolgonderzoek.

## Van interventiedoel naar veranderdoelen

### 2.1 Inleiding

Wat moet er in het statistiekonderwijs veranderen zodat studenten minder of geen problemen meer ondervinden met het leren, herinneren en toepassen van statistiek? In hoofdstuk 1 is als oplossing voorgesteld het vak statistiek te integreren met het vak methoden en technieken van onderzoek (M&T) in de context van psychologisch onderzoek. Het idee is dat studenten al onderzoekende de betekenis, de functie, de waarde en het nut van statistiek in psychologisch onderzoek zullen ervaren. De verwachting is dat zij hierdoor meer gemotiveerd raken zich de statistische kennis eigen te maken en dat hun attitude ten aanzien van de statistiek positiever wordt. Hierdoor zullen ook de kennis van en de vaardigheid in het toepassen van statistiek toenemen. Door integratie met het vak M&T ontstaan mogelijkheden de statistiekinhouden over meer cursussen te spreiden. Hierdoor is het niet meer nodig de complexe inhouden die hiërarchisch met elkaar samenhangen, compact in twee cursussen te persen.

In dit hoofdstuk wordt stap twee volgens Intervention Mapping uitgevoerd. Dit protocol schrijft in deze fase drie taken voor. In de eerste plaats moet de vraag worden beantwoord: 'Wat willen we precies dat de studenten leren?' Deze vraag staat centraal in paragraaf 2.2. Hier worden de gewenste leeruitkomsten verder uitgewerkt in concrete gedragsdoelen ('performance objectives'). Hoewel het ontwerp zich zal richten op de onderzoekscompetentie tot op het niveau van de bachelorthesis, zullen de gedragsdoelen in dit proefschriftonderzoek beperkt blijven tot de statistische aspecten van het onderzoek. Vervolgens moet worden nagegaan welke de determinanten zijn die de gewenste gedragingen kunnen bewerkstelligen. Intervention Mapping onderscheidt hierin persoonlijke en externe determinanten. De mogelijkheden om via de positionering van het vak statistiek de onderwijsleeromgeving dusdanig te veranderen dat de gedragsdoelen beter zullen worden bereikt worden besproken in paragraaf 2.3. Als laatste stap dienen de geselecteerde determinanten afgestemd te worden op de eerder geformuleerde gedragsdoelen. De hieruit ontstane veranderdoelen geven antwoord op de vraag: 'Wat zal er in de onderwijsleeromgeving moeten veranderen om te bereiken dat studenten de gedragsdoelen ook bereiken?' Het antwoord op die vraag wordt geformuleerd in paragraaf 2.4. Met de formulering van de veranderdoelen wordt het interventiedoel geconcretiseerd.



## 2.2 Bepaling van de gewenste leeruitkomsten: gedragsdoelen

Van afgestudeerde psychologen mag worden verwacht dat zij de competentie hebben om op wetenschappelijk verantwoorde wijze via empirisch onderzoek geldige en bruikbare kennis te genereren of te toetsen teneinde een vraag binnen hun vakgebied te kunnen beantwoorden. Deze verwachting is in overeenstemming met het academische ideaalbeeld van de praktiserende psycholoog als ‘scientist practitioner’ (Commissie Levelt, 2006; *Rapport Onderwijsvisitatie Psychologie*, 2007).

Dit vereist een integratie van drie theorieniveaus: inhoudelijke kennis (beschrijving van theorieën, gedrag, verschijnselen), methodologische theorieën (operationalisering van abstracte en complexe begrippen (i.c. psychologische constructen) naar onderzoekstechnische, meetbare begrippen en data-analytische theorieën (analyseren en interpreteren van onderzoeksgegevens) (Mellenbergh, 1980; Schuyten, 2001). In wetenschappelijk onderzoek sluiten deze drie niveaus naadloos op elkaar aan. Die integratie leidt tot vier gewenste leeruitkomsten of gedragsdoelen (‘performance objectives’).

Het eerste gedragsdoel betreft de *functionele* vakkennis van de statistiek. Functionele kennis impliceert volgens Biggs (2003, p. 42) dat studenten weten *wat* de statistische techniek inhoudt (declaratieve kennis), *hoe* deze techniek moet worden uitgevoerd (procedurele kennis) en *onder welke voorwaarden* die techniek mag worden toegepast (conditionele kennis). Statistische technieken die door studenten op het bachelorniveau ingezet moeten worden, reiken minimaal tot correlatie-, regressie-, variantieanalyse en chi-kwadraattoetsen.

Het tweede gedragsdoel betreft het kunnen samenwerken en communiceren met anderen over statistische analyses. Dit betekent dat ze statistische bevindingen van peers kunnen beoordelen en kunnen communiceren met experts (docenten) over meer geavanceerde statistische onderwerpen.

Het derde doel betreft het inzicht in de diverse wetenschappen. Studenten die academisch gevormd worden dienen zich bewust te zijn van de samenhang die er bestaat tussen de kennis van de psychologie en de wijze waarop die kennis tot stand is gekomen via disciplines als methodologie en statistiek. Zij transformeren hiermee van consument naar producent van kennis (Donald, 2002)

Het vierde gedragsdoel betreft de zelfstandigheid (of autonomie) van de student. Studenten moeten op zelfstandige wijze statistische technieken aan de juiste onderzoeksvraag kunnen koppelen en uitvoeren.

In tabel 2.1 zijn de gedragsdoelen waar nodig in subgedragingen gespecificeerd.

TABEL 2.1 Gedragsdoelen voor de statistische aspecten van de onderzoekskompetentie conform de cursusplannen van de opleiding Psychologie

<i>Gedragsdoelen Student</i>	<i>Bachelorstudent psychologie</i>
G.1 toont beheersing van functionele kennis van de statistiek bij onderzoeks- opdracht	<p>G1.1 geeft aan welke statistische technieken al dan niet mogelijk zijn bij een probleemstelling met een vraag naar samenhang (tussen variabelen) of een vraag naar verschil (tussen groepen)</p> <p>G1.2 geeft aan welke statistische technieken al dan niet mogelijk zijn bij de definiëring en de daarop volgende operationalisering van de begrippen en de relaties daartussen in de probleemstelling</p> <p>G1.3 geeft aan welke statistische technieken al dan niet mogelijk zijn bij de definiëring en de daaropvolgende operationalisering van de onderzochte groep(en) respondenten c.q. proefpersonen</p> <p>G1.4 selecteert de juiste statistische techniek gegeven een onderzoekssituatie</p> <p>G1.5 past de statistische techniek adequaat toe (m.b.v. SPSS) gegeven een onderzoekssituatie</p> <p>G1.6 interpreteert de resultaten van de statistische analyses correct gegeven een onderzoekssituatie</p> <p>G1.7 rapporteert de resultaten van de statistische analyses correct volgens de APA-regels gegeven een onderzoekssituatie</p>
G.2. beoordeelt vakliteratuur op statistische aspecten en communiceert hierover met vakgenoten	<p>G.2.1 beoordeelt wetenschappelijke artikelen op correcte toepassing van de statistiek</p> <p>G.2.2 wisselt epistemische interacties uit met collega-studenten</p> <p>G.2.3 communiceert met (statistiek)experts over geavanceerde onderwerpen</p> <p>G.2.4 documenteert resultaten van onderzoek adequaat met de bedoeling bij te dragen aan de kennisontwikkeling in het vakgebied en daarbuiten</p> <p>G.2.5 rapporteert bevindingen volgens de APA</p>
G.3. toont inzicht in de samenhang tussen de wetenschappen	G.3.1 geeft aan hoe in het onderzoeksproces de diverse disciplines samenhangen en van elkaar afhankelijk zijn
G.4. voert de statistische analyses zelfstandig uit	G.4.1 voert de analyses zelfstandig uit

### 2.3 Selectie van de determinanten

Onderzoek naar de effectiviteit van het onderwijs laat zien dat zowel de student als zijn leeromgeving belangrijke factoren zijn in het onderwijsleerproces (zie onder anderen Biggs, 2003). Zo blijkt bijvoorbeeld dat studentkenmerken als intelligentie, voorkennis en motivatie voor de studie, maar ook kenmerken van de opleiding zoals de wijze waarop leerdoelen, instructie en assessments op elkaar zijn afgestemd van invloed zijn op de wijze waarop de student zijn studie oppakt. In de terminologie van Intervention Mapping zijn studentkenmerken persoonlijke determinanten van gedrag en bevat de leeromgeving externe determinanten van gedrag.

In dit proefschrift wordt een determinantenmodel gehanteerd dat is gebaseerd op de theorie van zelfgereguleerd leren ('Self-Regulated Learning', Pintrich, 2004; Zusho & Pintrich, 2003) (zie figuur 2.1). Deze theorie stelt dat de interactie tussen studentkenmerken en leeromgevingskenmerken leidt tot leergelateerde processen. Studenten passen in de procesfase hun cognities en wijze van studeren (leerstrategie) aan, met als



gevolg een verbetering van het resultaat. Pintrich (2004) en Zuscho en Pintrich (2003) onderscheiden als cognities in hun theorie van zelfgereguleerd leren de ‘beweeg’-redenen om doelgericht actie te ondernemen en te onderhouden – zij noemen dat de motivatie (vanuit ‘movere’ dat beweging betekent) – met cognities in de vorm van leerstrategieën. Genoemde onderzoekers onderscheiden in hun theorie vier motivatie-componenten en vijf leerstrategieën (zie hierna in paragraaf 2.3.3). Uiteindelijk leiden de reacties van de studenten op de kenmerken van de leeromgeving tot een bepaald product zoals de kwaliteit en het niveau van functionele kennis van de statistiek (zie tabel 2.1). De vraag hoe motivatie het leren faciliteert en de prestaties beïnvloedt, is sinds enkele decennia uitgebreid onderwerp van onderzoek (zie Covington, 2000).

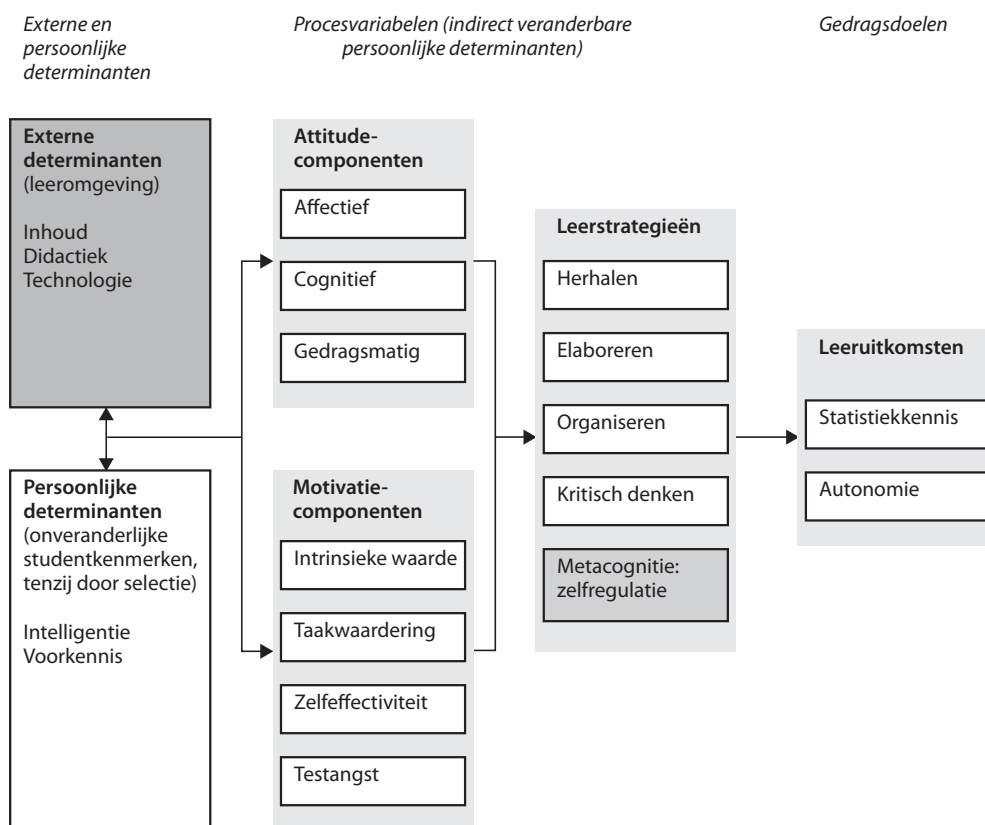
In de theorie van Pintrich (2004) en Zuscho en Pintrich (2003) ontbreekt de attitude ten aanzien van de statistiek. Ook dit psychologisch construct is in relatie tot het product (statistiekennis) uitgebreid voorwerp van onderzoek geweest en wordt ook als een van de verklaringen voor de problemen in het onderwijs beschouwd (Garfield & Ben-Zvi, 2007). Er zijn slechts enkele onderzoeken bekend waarin zowel motivatie als attitude simultaan in een model zijn opgenomen, maar uit deze onderzoeken wordt niet duidelijk wat de verschillen tussen beide constructen zijn of worden ze oppervlakkig uitgewerkt (Lalonde & Gardner, 1993; Nasser, 2004). In aanvulling op de theorie van Pintrich (2004) en van Zuscho en Pintrich (2003) wordt daarom attitude ten aanzien van de statistiek aan het determinantenmodel toegevoegd (zie figuur 2.1).

In het determinantenmodel worden motivatie, attitude en inzet van leerstrategieën als procesvariabelen opgevat die mediëren in de relatie tussen leeromgeving en leeruitkomsten. In termen van Intervention Mapping zijn deze procesvariabelen ook als persoonlijke determinanten te beschouwen die van belang zijn voor de realisatie van de gedragsdoelen. Deze psychologische constructen zijn echter niet op een directe wijze door de interventieontwikkelaar te veranderen. Ze worden geacht in positieve zin te veranderen door ingrepen in de externe omgeving.

Hoewel studentkenmerken als intelligentie, voorkennis, sekse en leeftijd ook van invloed zullen zijn op de leeruitkomsten, worden deze persoonlijke determinanten niet als niet beïnvloedbaar beschouwd (zijn daarom niet in grijs weergegeven). Reden daarvoor is dat veranderingen hierin alleen door middel van ‘selectie aan de poort’ kunnen plaatsvinden. Dit is bij de Open Universiteit Nederland niet mogelijk.

De interventie in dit proefschrift richt zich derhalve direct op de leeromgeving en dan meer specifiek op een wijziging in de positionering van het vak in het curriculum (2.3.1) en indirect op de genoemde procesvariabelen (2.3.2). De verwachting is daarmee bij te dragen aan een structurele oplossing van de problemen in het statistiekonderwijs. In de volgende paragrafen wordt uiteengezet hoe op basis van huidige inzichten over het leren ingrepen in bepaalde kenmerken van de leeromgeving tot de gewenste veranderingen in de gedragsdoelen zullen moeten leiden. De veranderdoelen, die

impliciet in de paragrafen (2.3.1) en (2.3.2) zijn terug te vinden worden gedetailleerd vermeld in de matrix van veranderdoelen (bijlage I.1).



FIGUUR 2.1 Determinantenmodel (naar Pintrich, 2004; Zusho & Pintrich, 2003)

### 2.3.1 Direct veranderbare externe determinanten

Moore (1997) wijst er op dat in het statistiekonderwijs gestreefd moet worden naar synergie tussen inhoud, didactiek en technologie. Door de positionering van de statistiek in het psychologiecurriculum te wijzigen, kunnen statistiekinhouden, didactiek en technologie beter worden toegesneden op de behoeften en capaciteiten van de student. De positionering bepaalt in hoge mate de vrijheidsgraden die statistiekdocenten en onderwijsontwerpers hebben om optimaal tegemoet te komen aan de wijze waarop studenten zich abstracte en complexe stof eigen maken. Hoewel inhoud, didactiek en technologie afzonderlijk zullen worden besproken, moet er worden opgemerkt dat deze determinanten nauw verweven zijn en elkaar wederzijds versterken en beïnvloeden. Zo zullen bijvoorbeeld onder het kopje inhoud ook didactische en technologische aspecten aan bod komen.

### Inhoud

Aan de inhoud van het statistiekonderwijs kan binnen de opleiding psychologie niet veel worden veranderd: de correlatie-, regressie- en variantieanalyse worden bij de meeste opleidingen beschouwd als een minimum. De meeste studenten ervaren deze inhouden als zeer abstract en complex en daarom moeilijk te leren. Pieters en Verschaffel definiëren leren als 'een constructief, cumulatief, doelgericht, gesitueerd, coöperatief en individueel verschillend proces van kennisverwerving' (in Verloop & Lowyck, 2003, p. 254; zie ook Shuell, 1988). De vraag is in welke mate het huidige statistiekonderwijs daaraan tegemoet komt. De genoemde aspecten van het leerproces worden hierna besproken.

*Constructief:* Onder constructief leren wordt verstaan dat studenten zelf in interactie met de omgeving waardevolle kennisinhouden en vaardigheden opbouwen. Over de wijze waarop mensen leren stelt de cognitieve psychologie dat lerenden netwerken van schema's en associaties samenstellen uit abstracte regels (rule-based) en voorbeelden die in hun geheel met hun algemene verbanden worden opgeslagen in het geheugen (exemplar-based), waardoor netwerken van inhoudelijk samenhangende begrippen tot stand komen (Anderson, 2000; Bransford, Brown, & Cocking, 2000). De cognitieve structuren bepalen niet alleen de opslag van kennis, maar ook het proces van opnemen en het terugvinden daarvan (Skemp, 1971). Norman en Regher (1996) en Anderson (2000) noemen de volgende cruciale factoren die de vorming van een netwerk in het geheugen sterk beïnvloeden. De eerste factor betreft de *betekenis* die aan een stimulus wordt toegekend: betekenisvolle informatie is veel gemakkelijker te onthouden en te reproduceren dan informatie die niet gekoppeld kan worden aan bestaande netwerken (Schmidt, 1997). De tweede factor omvat het aantal en de sterkte van de onderlinge associaties in het netwerk, bekend onder de naam *activatiespreiding*. Geheugensporen worden geactiveerd wanneer daarmee geassocieerde concepten actueel zijn. Zowel de activatie als de sterkte van de associaties spelen daarbij een belangrijke rol. De sterkte van associaties wordt vooral bepaald door praktiseren (Anderson, 2000).

Een expert onderscheidt zich van een leek door zijn uitgebreide, goed geordende, coherente en rijk vertakte netwerk van begrippen met concrete voorbeelden uit het domein in het langetermijngeheugen; hij of zij beschikt over efficiënte methoden om aan elkaar gerelateerde concepten terug te vinden en over procedures om deze informatie te gebruiken bij het oplossen van problemen. Zowel de hoeveelheid kennis als de wijze waarop die kennis is georganiseerd, verlenen de expert de status van iemand met 'diepgaand inzicht' (Bransford et al., 2000; Norman & Regher, 1996). Anderson (2000) onderscheidt drie fasen in leren in de ontwikkeling van beginner tot expert:

- *De cognitieve fase.* Deze fase bestaat voornamelijk uit het memoriseren van feiten, van declaratieve kennis door herhalen.
- *De associatieve fase.* In de tweede fase worden geleidelijk aan de misconcepties uit de eerste fase ontdekt en verwijderd en worden de verbanden (associaties) tussen de verschillende elementen, die nodig zijn voor een succesvolle prestatie, versterkt. Deze fase vormt de overgang van 'weten wat' naar 'weten hoe' en 'weten wanneer', door Biggs (2003) met 'functionele kennis' aangeduid. De procedures, die de out-

put vormen van de tweede fase, kunnen worden omschreven als productieregels (Singley & Anderson, 1989).

- *De autonome fase.* In deze fase worden de procedures steeds meer geautomatiseerd en sneller toegepast (Singley & Anderson, 1989).

Door praktiseren verbeteren accuraatheid en snelheid; het laatste staat bekend als tuning (Anderson, 2000).

Vanwege de beperkte mogelijkheden in het curriculum belemmert het huidige vakgerichte statistiekonderwijs de opbouw van een complexe samenhangende geheugenstructuur met veel en goed geordende informatie in het langetermijngeheugen. De betekenis van de statistiek ontgaat de meeste studenten indien deze als afzonderlijke deelvaardigheden, zoals het berekenen van een gemiddelde of het uitvoeren van een regressieanalyse worden getraind. Dit levert op korte termijn wel succes op, maar werkt de samenhang in het langetermijngeheugen tegen. Door statistiek in te bedden in voor hem of haar relevante psychologische onderzoekscontexten waarin statistische technieken orde kunnen scheppen in een onoverzichtelijke hoeveelheid data, wordt de student gemotiveerd zich af te vragen welke techniek adequaat is en hoe deze toe te passen.

*Cumulatief:* Een goed begrip van het vak statistiek vereist dat er een adequate cognitieve structuur wordt ontwikkeld die voldoende aanknopingspunten of ankerpunten bevat waaraan nieuwe kennis kan worden gekoppeld. Beperkte studietijd, dichtheid van het materiaal en de hiërarchische opbouw van de vakgerichte statistiek leiden ertoe dat begrippen op een lager hiërarchisch niveau onvoldoende worden gepraktiseerd om begrippen op een hoger niveau te kunnen begrijpen in de vorm van een consistent samenhangend cognitief netwerk. Het gevolg is dat de statistische leerstof nauwelijks geassocieerd wordt met psychologische concepten en in de vorm van gedissocieerde cognities noch beklijven noch op te halen zijn (Schmidt, 1997). Hierdoor wordt ook de grondslag voor *transfer* van het geleerde weggenomen. Onder transfer wordt de bekwaamheid verstaan om kennis en vaardigheden te kunnen inzetten buiten de beperkte contexten waarin zij zijn verworven. De kans op transfer wordt groter als de leeromgeving ervoor zorgt dat er een grote overeenkomst is in de taak en/of leerintentie in de trainingssituatie en de uiteindelijke toepassing- of testsituatie (Norman & Regher, 1996). Door de kennis en begrip van de statistiek en M&T vanaf het begin van de studie op te bouwen en daarna voortdurend via betekenisvolle problemen te activeren, toe te passen en uit te breiden ('continue progressieve recontextualisatie' Van der Sanden, 2004) zal deze langer beklijven en neemt de kans op transfer toe (Lovett & Greenhouse, 2000; Van Merriënboer, 1997). Door voor statistiek een langere leer- en toepassingslijn te creëren hebben studenten meer mogelijkheden een goed geordend cognitief netwerk op te bouwen.

Een ander probleem dat zich aandient bij de behandeling van statistiek in compacte cursusvorm is dat studenten zich in korte tijd een veelheid aan nieuwe statistische begrippen en procedures dienen eigen te maken. Hierdoor neemt het risico van in-

effectieve cognitieve belasting ('extraneous cognitive load') toe en blijft er minder ruimte over voor effectieve cognitieve belasting ('germane cognitive load' (Van Merriënboer & Sweller, 2005; Wouters, 2007). Door het inpassen van de statistiek in een langere leer- en toepassingslijn wordt de kans op ineffektieve cognitieve belasting verminderd

*Doelgericht:* Een leerproces is het meest effectief wanneer studenten hun eigen doelen nastreven. Een uitdaging voor een opleiding Psychologie is volgens Donald (2002) '...first by linking scientific knowledge with students' interests by means of powerful metaphors – mind as reflex machine or knower – and second, by introducing students to the process of research that defines the field [...] the second challenge is to provide students with positive experiences in learning to do research – for example, exercises that help them realize they already know a considerable amount of the steps of research process so that they are more confident in dealing with design and statistical issues.' (p. 166).

Binnen het geïsoleerd gepresenteerde statistiekonderwijs leert de student puur statistische vraagstukken oplossen met een geleerd statistisch principe. Wanneer de student echter in de eindfase van zijn/haar studie bij de onderzoeksthesis voor een data-analytisch probleem wordt gesteld, wordt er van hem of haar verwacht dat hij of zij kan uitmaken welk concept relevant is voor het actuele psychologische vraagstuk. Als men wil bereiken dat studenten in een bepaalde situatie de juiste statistische techniek kiezen, dan moeten ze tijdens de instructie ook praktiseren in het kiezen van de juiste statistische maat. Onderzoek naar redeneerprocessen maakt duidelijk dat mensen hun redeneringen vooral baseren op specifieke contextuele informatie, die tijdens het leren is opgeslagen (Anderson, 2000; Norman & Regehr, 1996).

*Gesitueerd:* Het leren is gesitueerd omdat het plaatsvindt in een sociale en materiële context. Veel leren gaat gepaard met het kennisnemen van gewoonten, gebruiken, kennis en vaardigheden die voor het adequaat functioneren als lid van een sociaal-culturele gemeenschap noodzakelijk zijn. Kennis die los van elke concrete situatie en ervaring ontstaat, beklijft niet en is daardoor moeilijk toepasbaar. Het is daarom raadzaam om de link te leggen naar psychologische probleemstellingen. Volgens de theorie van het contextgebonden denken ('situated cognition': Brown, Collins, & Duguid, 1989) zijn denken en leren geen processen die zich geïsoleerd in het hoofd afspelen en die resulteren in de opbouw van abstracte voorstellingen. Het zijn 'gedeelde' activiteiten, waarbij de denk- of leeractiviteit wordt verdeeld over de student zelf, hulpmiddelen (studietaken, bronnen) en andere personen (begeleiders en medestudenten) die in de competentiegerichte leeromgeving aanwezig zijn.

*Coöperatief:* Het coöperatieve aspect is het leren van en met anderen. Diverse studies hebben aangetoond dat samenwerkend leren superieur is aan individueel leren en aan competitie (Johnson, Johnson, & Stanne, 2000). Uit onderzoek blijkt dat het denken wordt gestimuleerd als men verschillende perspectieven ervaart en daar-

over van gedachten wisselt (Tomic & Span, 1993). Daarbij is het belangrijk dat de ‘natuurlijke’ aanleiding voor de discussie een probleem in een context is en dat de interactie gebeurt tussen de studenten onderling en uiteraard ook tussen de docent en de student.

Tomic en Span (1993) vatten de betekenis van ‘interactie’ als volgt samen:

- Door zijn ideeën onder woorden te brengen krijgt de student meer greep op zijn denken.
- De docent van zijn kant krijgt ook meer zicht op de denkprocessen van de studenten.
- Door met elkaar te discussiëren worden studenten gedwongen niet alleen zo scherp mogelijk te formuleren, maar ook te luisteren, mee te denken en zich in te leven in de bedoeling van de ander.

Ohlsson (1996) heeft het over ‘epistemische’ interactie en omschrijft deze als een bundel taken zoals: beschrijven, verklaren, voorspellen, argumenteren, kritiek geven en evalueren, uiteenzetten en elaboreren en definiëren. Deze epistemische interactie, waarvan feedback een essentieel onderdeel vormt, leidt tot hogere orde denken dat op zijn beurt weer leidt tot diepere verwerking en langer vasthouden van leerstof (Felder & Brent, 2001; McKeachie & Hofer, 2001). Bovendien geldt epistemische interactie als een belangrijk element binnen actief leren (‘active learning’). Volgens dit concept krijgt de student, via opdrachten, taken, groepswork, een centrale rol waarbij de verantwoordelijkheid van de begeleider verschuift van docent die lesgeeft (volgens het traditionele model) naar een begeleider die waakt over het leerproces en dit faciliteert.

Het vakgerichte statistiekonderwijs is vanwege zijn aard deductief en individualistisch georiënteerd en laat daardoor weinig ruimte over voor discussie. Voor het doen van onderzoek met ‘ill-structured’-problemen die een meer open einde hebben voor oplossingen en daardoor geschikter zijn voor epistemische interactie, kan het samenwerken in groepen een diepgaander verwerking van de leerstof tot gevolg hebben. Hoewel de Open Universiteit Nederland als instelling voor afstandsonderwijs geen campus heeft waar studenten elkaar kunnen treffen, biedt internet in de vorm van ‘communities of learners’ of ‘communities-of-practice’ mogelijkheden tot het instellen van virtuele groepen. Waar mogelijk moet het onderwijs dus voorzien in opdrachten die tot samenwerking leiden. Wel zal rekening moeten worden gehouden met de verschillende groepen studenten die hierin zijn te onderscheiden. Niet alle studenten die hebben gekozen voor de Open Universiteit Nederland zullen dit kunnen of willen, maar het streven is om het te stimuleren. Er zijn studenten die het aangenaam vinden om met elkaar te kunnen samenwerken. Voor afstandsstudenten kan een elektronische leeromgeving uitkomst bieden. Een andere groep wil wel passief deelnemen, dus door berichten te lezen in de nieuwsgroepen en zich zodoende een beeld te vormen over de interacties van anderen (vergelijk sites waarin consumenten ervaringen kunnen uitwisselen). Door de discussies van anderen te vormen vindt de ‘passieve’ student vaak een oplossing voor zijn eigen problemen. Weer een andere



groep zal wel willen samenwerken, maar heeft daar verder geen uitgesproken mening over. En de laatste groep kiest voor de Open Universiteit Nederland omdat ze dan geheel in eigen tempo kunnen studeren.

*Individueel:* Studenten van de Open Universiteit Nederland zijn niet gebonden aan klassen of groepen. Bovendien kunnen zij op elk gewenst tijdstip beginnen met hun studie met een cursus naar keuze. De opleiding is nagenoeg geheel geïndividualiseerd. Het individuele aspect betreft de kenmerken die het leerproces van de student uniek maken. Hierbij sluit personalisatie van studietaken aan, waardoor beter kan worden ingespeeld op bijvoorbeeld voorkennisverschillen en verschillen in interessegebieden (bijvoorbeeld arbeids- en organisatiepsychologische versus gezondheidspsychologische casussen, eigen onderzoeksvragen).

#### Samengevat

Betekenisverlening en fasen in het leren stroken niet met de wijze waarop het vakgerichte statistiekonderwijs in de vorm van enkele compacte servicecursussen is georganiseerd. Het abstracte karakter van de statistiek met de sterk hiërarchische en deductieve opbouw van de leerstof sluit slecht aan bij de verschillende behoeften en strategieën van psychologiestudenten. De leerfasen zoals die door Anderson (2000) zijn onderscheiden, verlangen een langlopende leer- en toepassingslijn. Door de verwerving van de statistiek te integreren met methoden en technieken van onderzoek ontstaat de mogelijkheid zowel de studenten de functie van statistiek binnen een onderzoekscontext als betekenisvol te laten ervaren als een langlopende leer- en toepassingslijn te creëren.

#### Didactiek

Een meer activerende onderwijsbenadering vereist dat studenten niet alleen *lezen* over onderzoek, maar het ook *doen*. Voor het verwerven van competenties onderscheiden Kirschner, Van Vilsteren, Hummel en Wigman (1997) in een onderwijsleeromgeving de taakomgeving en de kennisomgeving. In de *taakomgeving* stellen docenten onderzoeksoopdrachten. Studenten nemen daarin de rol aan van onderzoeker. In deze rol dienen zij zich af te vragen wat de typische eisen zijn die aan een onderzoeker wordt gesteld. Door de onderzoeksoopdracht te situeren in een authentieke of realistische context, laat de competentie zich onderscheiden van vaardigheid. De complexiteit van de taak wordt grotendeels bepaald door de context: de mate van informatie, de mate van tijd om het onderzoek uit te voeren, enzovoorts. De context maakt de taak ook complexer dan men gewend is in een meer schoolse situatie, waarin de variabelen veelal variëren binnen het specifieke kennisdomein en er geen sprake is van allerlei randvoorwaarden en versturende invloeden. Afhankelijk van de onderzoekssituatie moeten kennis en vaardigheden steeds weer in andere combinaties en op een gecoördineerde wijze worden toegepast en uitgevoerd. Aan de onderzoekstaak dienen criteria te worden gekoppeld die aangeven waaraan de beantwoording van de onderzoeksvraag moet voldoen en hoe de student kan worden gekwalificeerd als bijvoorbeeld een slechte, matige, goede of uitmuntende onderzoeker. Heeft de student

de onderzoeksituatie afdoende geanalyseerd? Is de argumentatie van het antwoord plausibel? En van belang voor het statistisch denken: passen de geselecteerde analysetechnieken bij de keuzen die gedaan zijn in het onderzoeksproces en worden ze op correcte wijze toegepast?

Gezien de complexiteit van onderzoeksprocessen en de daarmee gepaard gaande verwerving van complexe vaardigheden (bijvoorbeeld bepaalde statistische technieken) is het duidelijk dat de onderzoekscompetentie niet in één enkele onderzoeksrondte kan worden bereikt. Docenten zullen moeten voorzien in een reeks van onderzoeksronden waarvan de complexiteit gaandeweg toeneemt. Figuur 1.1 met de competentiecirkel illustreert deze opklimmende reeks van onderzoeksronden die tot het niveau van de bachelorthesis reiken.

Daarnaast zijn er ontwikkelingen te onderkennen waarmee het statistiekonderwijs zijn voordeel kan doen. Gedoeld wordt op het huidige onderzoek op het terrein van het statistiekonderwijs dat allerlei initiatieven laat zien die het onderwijsleerproces kunnen bevorderen. Docenten experimenteren met een minder formele behandeling van de statistiek ('meer data, minder formules', Garfield, delMas, & Chance, 2007; Moore, 1997; Noll, 2007; Schwartz et al., 2007; Zieffler, 2006), speciale trainingen (Quilici & Mayer, 2002), diverse vormen van feedback (Lovett, 2001); gebruik van technologie (bijvoorbeeld Moonen, 1978; Noll, 2007), simulaties (Lane & Peres, 2006; Lane & Tang, 2000; Mills, 2004), online en webgebaseerd onderwijs in plaats van klassikaal onderwijs (Utts, 2003; Ward, 2004) en samenwerkend leren (delMas, Garfield, & Chance, 1999; Garfield, 1993; Giraud, 1997; Keeler & Steinhorst, 1995; Magel, 1998).

In de *kennisomgeving* vindt de student de materialen en bronnen om de studietaak te kunnen uitvoeren (zie hierna onder ad 3). Het onderscheid tussen een taak- en een kennisomgeving stelt de onderwijsontwerper in staat in de loop van de tijd state-of-the-art kennis toe te voegen of ernaar te verwijzen zonder dat daardoor de uit te voeren taak hoeft te worden veranderd. Het betekent alleen dat van studenten wordt verwacht dat ze een andere, meer adequate techniek of een nieuw inzicht toepassen in de taakomgeving (Kirschner, Van Vilsteren, Hummel, & Wigman, 1997).

Met de invoering van de competentiegericht onderwijs, moet ook de beoordeling worden aangepast. Dit is een kwestie van congruentie of 'alignment' (Biggs, 1996, 1999; Entwistle, 2000; Elshout-Mohr & Oostdam, 2001). Leerstrategieën blijken nauw samen te hangen met de uiteindelijke leerresultaten en dat betekent dat assessments niet los kunnen worden gezien van onderwijs (Biggs, 1996, 1999; Alderson & Wall, 1993). De manier van tentamens afnemen of de wijze van toetsing heeft een 'washback'-effect op de gehanteerde leerstrategie (Alderson & Wall, 1993; Baartman, Bastiaens, & Kirschner, 2004; Biggs, 2003; Jochems, Merriënboer, & Koper, 2004). Het gebruik van meerkeuzetoetsen bij het vakgerichte statistiekonderwijs met onafhankelijke items is in veel opleidingen gebruikelijk. Voor het beoordelen of iemand kan onderzoeken is deze manier van toetsen niet de meest adequate. Nagegaan moet worden in hoeverre



andere toetsingsmethoden beter op één lijn met de leerdoelen en de instructie in overeenstemming worden gebracht. Een punt van overweging zijn wel de begeleidingskosten die hiermee gemoeid zijn.

### *Technologie*

Het afstandskarakter van het OUNL-onderwijs impliceert dat studenten hun onderwijsmaterialen op hun eigen werkplek afgeleverd krijgen. De begeleiding geschiedt eveneens op afstand. De cursussen statistiek kennen enkele – facultatieve – groepsbijeenkomsten in een van de studiecentra in het land onder leiding van een docent. Met de opkomst van ICT en het gebruik van allerlei software is de werkplek van de student thuis achter de computer. Deze laatste kan ook ingezet worden als een medium waarmee opdrachten, taken, bronnen, materialen kunnen worden verspreid, maar ook als een middel voor asynchrone (e-mail, discussiegroepen) als synchrone (online) communicatie. De elektronische leeromgeving van de OUNL (Studienet) die sinds 1998 in de lucht is, kan studenten de mogelijkheid bieden een persoonlijke werkplek op te bouwen en van daaruit nieuws en dergelijke over de opleiding te verkrijgen, extra door de faculteit beschikbaar gestelde informatiebronnen te downloaden en via nieuws- en discussiegroepen plaats- en tijdonafhankelijk te kunnen communiceren met medestudenten en docenten. Deze internetsites maken afstandsleren flexibeler, aantrekkelijker en persoonlijker dan voorheen.

### 2.3.2 *Indirect veranderbare persoonlijke determinanten: procesvariabelen*

In het determinantenmodel worden attitude, motivatie en leerstrategieën als procesvariabelen beschouwd. Dat wil niet zeggen dat deze variabelen alleen als procesvariabelen het proces tussen leeromgeving (in interactie met de studentkenmerken) en product mediëren. In werkelijkheid zullen in een dynamisch systeem genoemde variabelen óók als moderatorvariabelen kunnen optreden (Judd, Kenny, & McClelland, 2001; Little, Card, Bovaird, Preacher & Crandall, 2007; MacKinnon, Fairchild & Fritz, 2007).

### *Attitude*

De attitude van studenten ten opzichte van statistiek wordt door statistiekdocenten vaak genoemd als een van de oorzaken van de problemen in het statistiekonderwijs (zie onder andere Gal, Ginsburg, & Schau, 1997; Gordon, 1999, 2004; Schau, Stevens, Dauphinee, & Del Vecchio, 1995; Sorge & Schau, 2002; Wisenbaker, Scott & Nasser, 1999, 2002). Een van de eerste onderzoekers die zich met attitudes bezig hield, definieerde attitude als ‘en mentale toestand, georganiseerd door ervaring die richting geeft aan de reacties van een individu met betrekking tot een object, situatie of gedrag’ (Allport, 1935).

Eagly en Chaiken (1993) brengen het evaluatieve karakter van de attitude in hun definities naar voren. Zij definiëren attitude als ‘een psychologische tendens die tot uitdrukking komt in een min of meer positieve of negatieve evaluatie van een object’ en als ‘een evaluatieve tendens die bemiddelt tussen bepaalde klassen van stimuli en

bepaalde klassen van responsen'. Attitudes worden omschreven als een amalgaam van affectieve, cognitieve en gedragsmatige associaties, die ingebed zijn in de structuren van de meer uitgebreide cognitieve schema's (Eagly & Chaiken, 1993).

Een attitude geeft dus een mentale toestand weer die ertoe leidt dat een individu op een consequente manier reageert op een bepaald verschijnsel in de vorm van goedkeuren of afkeuren, opzoeken of mijden dan wel waarderen of niet-waarderen. Voorts zijn attitudes gevormd door ervaringen, zowel direct ('Wat is de betekenis van de statistiek met al die cijfertjes voor mijn psychologiestudie? Psychologie gaat toch over mensen?') als indirect (Medestudenten: 'statistiek is moeilijk' en 'Je hebt het later als psycholoog toch niet meer nodig'). Blijkbaar zijn attitudes ook gebaseerd op verwachte consequenties ofwel de waargenomen functie van de statistiek voor de psychologie. Eagly en Chaiken (1993) omschrijven attitudes ook als gesocialiseerde meningen, gevoelens en gedrag. Het socialisatieproces begint al bij de geboorte. Daarom zijn attitudes doorgaans diep geworteld, relatief duurzaam en stabiel, semibewust en bijna automatisch in hun uitwerking.

Bovenstaande samenvattend wordt in dit proefschrift de attitude ten aanzien van de statistiek gedefinieerd als de stabiele, evaluatieve tendens ten aanzien van statistiek, gebaseerd op directe (ofwel eigen) en indirecte (ofwel andermans) ervaringen en verwachte consequenties van de bestudering van statistiek, die richting geeft aan de wijze waarop statistiek wordt bestudeerd.

Zoals eerder vermeld is de attitude bij psychologiestudenten ten aanzien van statistiek bepaald niet positief (Gordon, 1999, 2004). Ook blijken studenten volgens Verschuren (2002) en Gelso (2006) een vertekend beeld te verwerven van onderzoek doen door het apart aanbieden van statistiek en methoden van onderzoek. Met de beoogde interventie wordt geprobeerd het negatieve affect ten aanzien van statistiek te verminderen, belangstelling en interesse voor onderzoek te wekken en daardoor de cognitieve en affectieve betrokkenheid bij het leren te verhogen.

### *Motivatie*

Pintrich en Schunk (2002) definiëren motivatie als 'het proces waarbij doelgerichte actie in gang wordt gezet en in stand wordt gehouden'. Pintrich (2004) en Zusho en Pintrich (2003) verbinden in hun model van zelfgereguleerd leren motivatie met een bepaalde inzet van leerstrategieën. In hun model onderscheiden zij vier motivatie-componenten: intrinsieke waarde, taakwaardering, zelfeffectiviteit en testangst. Deze vier componenten zullen nu kort worden toegelicht.

De component *Intrinsieke Waarde* geeft aan of men zich betrokken voelt bij een leertaak omdat men deze als uitdagend ervaart, de nieuwsgierigheid prikkelend of omdat deze leidt tot vaardigheden en competenties die op beheersing ('mastery' en zelfregulatie) wijzen (Pintrich & De Groot, 1990; Vansteenkiste, Simons, Lens, & Soenens, 2004a, 2004b; Lewalter & Krapp, 2004). Onderzoek wijst erop dat de intrinsieke waarde van taken zeer gevoelig is voor contextuele invloeden (Eccles, 1983), zoals de karakteristieken van een taak of de discipline (Middleton & Spanias, 1999). Lewalter en

Krapp (2004) constateren dat vooral taken die uitdagen tot actieve participatie van de student in staat zijn de interesse van de student te wekken. Martens, Gulikers, & Bastiaens (2004) en Van Merriënboer (1997) stellen dat vooral authentieke taken de interesse wekken van de student.

Leertaken prikkelen tevens de motivatiecomponent *Taakwaardering* (Eccles & Wigfield, 2002; Pintrich & Schunk, 2002). Taakwaardering is de evaluatie van het belang en het nut van een taak (Pintrich, Smith, Garcia, & McKeachie, 1991, 1993; Linnenbrink & Pintrich, 2003). Bij deze vorm van motivatie is de activiteit instrumenteel, een middel om het doel te bereiken (Pintrich & Schunk, 2002).

*Zelfeffectiviteit* (self-efficacy) wordt gedefinieerd als 'de gepercipieerde bekwaamheid om te leren of te handelen op een bepaald niveau' (Bandura, 1993; 1997) en als 'evaluaties van mensen over hun eigen capaciteit om doelgericht gedrag te organiseren en uit te voeren' (Bandura 1986, p. 391, in Pintrich & Schunk, 2002). Hiermee verbindt Bandura (1986) zelfeffectiviteit met doelen. Zelfeffectiviteit heeft invloed op de keuze van acties, inspanningen en doorzettingsvermogen. Mensen met een lage zelfeffectiviteit voor een taak kunnen zich bedienen van vermijdingsgedrag (vermijdingsdoelen, Linnenbrink & Pintrich, 2003; Pintrich, 2002). Bij een hoge zelfeffectiviteit is het waarschijnlijker dat men aan de taak deelneemt omdat men denkt in staat te zijn om de taak goed uit te kunnen voeren. Bij de confrontatie met problemen pakken zelfeffectieve studenten harder aan en zetten ze langer door dan studenten met een lage zelfeffectiviteit die aan hun eigen capaciteiten twijfelen. Zelfeffectiviteit bepaalt in belangrijke mate de gedragsmatige betrokkenheid bij het leren, die zichtbaar en meetbaar is in de vorm van inspanningen, doorzetten en het zoeken van hulp om er iets van te leren (Linnenbrink & Pintrich, 2003). Bij een lage zelfeffectiviteit is men veel minder geneigd om hulp te zoeken omdat men denkt dat anderen hen dan dom zullen vinden (zelfbeeld). Bij een hoge zelfeffectiviteit heeft men daar geen last van (Karabenick & Knapp, 1991; Newman, 1990; Ryan & Pintrich, 1997, in Linnenbrink & Pintrich, 2003). Correlatiestudies bevestigen dat zelfeffectiviteit een predictor is van gedragsmatige inspanningen en doorzetten (Schunk, 1989, 1991, Pintrich & Schunk, 2002). Experimenten met feedback om zelfeffectiviteit te verhogen (Jackson, 2002, Schunk, 1989, 1991, Pintrich & Schunk, 2002) tonen aan dat zelfeffectiviteit te beïnvloeden is door middel van positief opbouwende feedback. Zelfeffectiviteit komt tot stand door informatie uit het actuele gedrag, door observaties, vormen van overtuiging, feedback en sociale-vergelijkingsprocessen (Bandura, 1986; Vrugt, 1994). De situationele gevoeligheid van zelfeffectiviteit betekent ook dat één en dezelfde student bijvoorbeeld een hoge zelfeffectiviteit kan hebben op het gebied van psychologische theorieën en een lage zelfeffectiviteit op het gebied van psychologisch onderzoek doen of vice versa. Zelfeffectiviteit blijkt een krachtige predictor van leren en prestaties, zelfs als er gecontroleerd wordt op voorkennis en cognitieve vaardigheden (Hommes, 2006; Schunk, 1982, Vrugt, 1994). Schouwenburg (1998) stelt vast dat de gemiddelde correlatie tussen zelfeffectiviteit en leerresultaat doorgaans .35 bedraagt. Ook het experiment van Jackson (2002) bevestigt dat zelfeffectiviteit significant verbonden is met tentamenresultaten.

*Testangst* heeft te maken met de zorgen die men zich maakt over de mogelijke negatieve consequenties van de prestaties die geleverd worden tijdens een tentamen of een vergelijkbare beoordelingssituatie (Onwuegbuzie & Wilson, 2003; Zeidner, 1991). Het concept is verbonden met vermijdingsdoelen (Linnenbrink & Pintrich, 2002) en speelt een substantiële rol in het leerproces, vooral in domeinen als wiskunde en statistiek (Baloglu, 2001; Gal & Ginsburg, 1994; Garcia & McKeachie, 2005; Schunk, 2005). Testangst kan leren en presteren belemmeren of bevorderen. Zowel te lage als te hoge testangst kunnen belemmerend werken op leerprestaties (Blom, Severiens, Broekkamp, & Hoek 2005).

### *Leerstrategieën*

In het theoretisch model wordt zelfgereguleerd leren beschouwd als een leerpatroon dat is samengesteld uit een samenspel van het herhalen, organiseren, elaboreren (uitbreiden en bewerken) van de leerstof, kritisch denken en metacognities, ingebed in een structuur van motivatiecomponenten en hulpstrategieën (Blom et al., 2005; Pintrich et al., 1991, 1993; Pintrich 1994).

*Herhalen* (memoriseren) is een leerstrategie die vooral geschikt is voor relatief eenvoudige taken (Pintrich et al., 1991), voor steeds terugkerende routinetaken (Van Merriënboer, 1997) en voor het oproepen van informatie uit het werkgeheugen. De strategie is niet effectief bij het tot stand brengen van verbanden tussen informatie of bij de integratie van informatie met voorkennis.

Door te *organiseren* selecteert de student kernconcepten in de leerstof en brengt structuur in de informatie aan door deze te clusteren en te omlijnen. Het is een activiteit die inspanning vergt en verbonden is met gedragsmatige betrokkenheid bij de taak dat moet leiden tot betere leerresultaten (Pintrich et al., 1991, 1993).

*Elaboreren* omvat strategieën als parafraseren, samenvatten, het gebruik van analogieën en het maken van aantekeningen als ondersteuning bij het integreren van nieuwe informatie in voorkennis (Pintrich et al., 1991, 1993) evenals het verrichten van extra inspanningen bij het zoeken naar aanvullende leerbronnen (Biggs, 2003). Het helpt bij het opslaan van informatie in het langetermijngeheugen en brengt verbindingen tot stand tussen voorkennis en de verschillende onderwerpen die bestudeerd worden (Anderson, 2000; Biggs & Collis, 1982).

*Kritisch denken* wordt door Pintrich et al. (1991, 1993) gedefinieerd als de mate waarin studenten hun kennis toepassen in een variëteit aan situaties bij het oplossen van problemen, het nemen van beslissingen en het kritisch evalueren van de situaties aan de hand van criteria uit de betrokken discipline (Pintrich et al., 1991, 1993; Donald, 2002). Kritisch denken omvat naast componenten van logisch denken, problemen oplossen en abstraheren (Donald, 2002) ook autonomie in het denken en handelen in combinatie met het richten van de aandacht (Brown & Ryan, 2003; Erickson & Kruschke, 1998). Kritisch denken wordt onder andere gestimuleerd door groepsdiscussies. Discussie

is belangrijk omdat studenten dan gestimuleerd worden hun gedachten en argumenten te verwoorden, wat vervolgens helpt om die gedachten helder te krijgen en te structureren. Ohlson (1996) onderscheidt zeven vormen van dergelijke ‘epistemische’ interacties: beschrijven, verklaren, voorspellen, argumenteren, kritiseren/evalueren, uiteenzetten/elaboreren en definiëren (p. 51). Door problemen en ideeën uit te leggen aan anderen, ontstaat een diepgaander begrip.

*Metacognitieve zelfregulatie* is een coördinerende strategie, waarmee het leerproces wordt gepland, gecontroleerd (monitoren), richting gegeven, geëvalueerd en volhard. Het plannen omvat het stellen van doelen (Pintrich, 1994) en het analyseren van de taak om voorkennis te activeren, waardoor het gemakkelijker wordt om de leerstof te organiseren en te begrijpen (Pintrich et al., 1991, 1993). Monitoren omvat het sturen en richten van de aandacht en het zichzelf testen door het stellen van vragen. Dit ondersteunt het begrijpen en het doorgronden van de leerstof en de integratie met voorkennis (Pintrich et al., 1991, 1993). Regulatie betreft het afstemmen en continu evalueren van de eigen cognitieve activiteiten. Regulerende activiteiten verbeteren het leerproces omdat ze de student ondersteunen bij het controleren en het corrigeren van het leergedrag bij het uitvoeren van de leertaak (Pintrich et al., 1991, 1993). Inspanningscontrole betreft de vaardigheid om de eigen inspanningen en aandacht te richten en bij te sturen in situaties waarin men afgeleid dreigt te worden of te maken heeft met taken die men als saai ervaart (Pintrich et al., 1991). Inspanningscontrole is een vorm van zelfmanagement om doelen te volbrengen, ook in lastige situaties. Inspanningscontrole levert een belangrijke bijdrage aan academisch succes omdat het volharding in het gebruik van leerstrategieën waarborgt en reguleert (Pintrich et al., 1991, 1993).

Leerstrategieën zijn – zoals eerder gesteld – niet verbonden met de eigenschappen van de student (Schouwenburg, 1998) maar strategieën, waar een student gebruik van maakt, afhankelijk van de mate waarin het geboden onderwijs hem of haar daartoe uitdaagt (Biggs, 2003). Een leerstrategie is beweeglijker en flexibeler dan een leerstijl, want een leerstrategie omvat zowel een bewuste keuze van de student (zelfregulatie en keuze voor de tactieken) als invloed van de omgeving (de leertaak) voor die keuze. Combinaties van leerstrategieën en keuzes van leerstrategieën bepalen de variatie in leeractiviteiten, in tegenstelling tot leerstijlen, die de consistentie bepalen in meerdere situaties (Ramsden, 1992). Het kiezen van leerstrategieën sluit ook aan bij de volgens Ryan en Deci (2000) inherent menselijke psychologische basisbehoefte aan autonomie.

## 2.4 **Matrix van veranderdoelen**

De laatste taak van stap 2 van Intervention Mapping bestaat uit het combineren van de gedragsdoelen met de determinanten. Hiertoe worden de gedragsdoelen in een matrix gekruist met de relevante determinanten. In de cellen van deze matrix worden de veranderdoelen geformuleerd. Een veranderingsdoel is het antwoord op de vraag:

‘Wat willen we precies dat de studenten leren van de onderzoekscompetentie?’ en voor de externe determinanten: ‘Wat willen we precies dat er in de leeromgeving verandert, dusdanig dat de persoonlijke determinanten (attituden en motivaties) mee veranderen?’ In bijlage I.1 is de matrix met veranderdoelen opgenomen. De determinanten zijn in de top geplaatst en de gedragsdoelen in de linkerkolom. Met de matrix van veranderdoelen is stap 2 in Intervention Mapping gezet en kan worden overgegaan op de volgende stap. Deze houdt in dat bij de veranderdoelen instructieaanwijzingen uit de onderwijskunde worden geselecteerd waarmee deze doelen gerealiseerd kunnen worden. Dit is het onderwerp van hoofdstuk 3. Door deze instructieaanwijzingen vervolgens langs de rode draad van ‘het doen van psychologisch onderzoek’ te vlechten ontstaat een blauwdruk voor het onderwijsontwerp waarmee de interventie kan worden uitgevoerd (zie hoofdstuk 4), waarmee naar verwachting de problemen in dat onderwijs worden weggenomen.



## Van veranderdoelen naar instructieaanwijzingen

### 3.1 Inleiding

In hoofdstuk 2 is het globale doel van de interventie vertaald in concrete veranderdoelen die zijn toegesneden op veranderingen die in de leeromgeving van het statistiekonderwijs moeten plaatsvinden. In dit hoofdstuk wordt beschreven welke theoretische methodieken en praktijken (in de terminologie van Intervention Mapping: ‘theory-informed intervention methods’ en ‘practical strategies’) uit het onderwijsveld bijdragen tot het verwezenlijken van de veranderdoelen. Deze inventarisatie heeft geleid tot een selectie van methodieken en praktijken uit de cognitivistische en constructivistische stromingen in de onderwijskunde, in het bijzonder de ‘Cognitieve Load Theory’, ‘Meaningful Learning’ en het ‘Cognitive Apprenticeship Instruction Model’. Paragraaf 3.2 schetst de drie methodieken en praktijken en de daaruit afgeleide instructieaanwijzingen. In paragraaf 3.3 wordt de ‘Theory of the Research Training Environment’ toegelicht waarin Gelso (2006) een tiental componenten uit de wereld van (klinisch-psychologisch) onderzoek benoemt die vooral de cultuur rondom het (leren) doen van onderzoek beschrijft en de effecten op attitude en motivatie. Vijftien jaar na het poneren van zijn theorie is Gelso nagegaan in hoeverre de door hem genoemde cultuurcomponenten evident zijn gebleken. In paragraaf 3.4 worden alle instructieaanwijzingen gebundeld in een matrix. Deze matrix dient als basis voor het ontwerp waarin de instructieaanwijzingen simultaan worden toegepast.

### 3.2 Theoretische methodieken, praktijken en instructieaanwijzingen

Deze paragraaf haalt enkele theorieën aan uit de cognitivistische en de constructivistische stromingen in de onderwijskunde die relevant zijn voor het verwerven van complexe vaardigheden. Vanuit deze theorieën dienen instructieaanwijzingen gedestilleerd te worden die de vernaderdoelen uit hoofdstuk 2 in het onderwijs vorm kunnen geven. De theorieën zijn de Cognitive Load Theory (paragraaf 3.2.1), Meaningful Learning (paragraaf 3.2.2) en het Cognitive Apprenticeship Instruction Model (CAIM) (paragraaf 3.2.3).

#### 3.2.1 *Cognitive Load Theory*

De meeste cognitivistische modellen gaan ervan uit dat het werkgeheugen een beperkte capaciteit heeft. De *Cognitive Load Theory* (Van Merriënboer & Sweller, 2005) die voortbouwt op het werk van Miller (1956), stelt dat mensen optimaal leren wanneer de belasting van het werkgeheugen minimaal wordt gehouden om zo de verbinding met



de schematische structuren van het onbeperkte langetermijngeheugen te vergemakkelijken. Miller (1956) stelt dat het werkgeheugen in normale omstandigheden beperkt is tot zeven 'chunks' informatie; bij complex leren zelfs tot niet meer dan twee à vier chunks (Van Merriënboer & Sweller, 2005). Een overbelasting van het werkgeheugen leidt tot verwarring en vergeten informatie. Nieuwe informatie in het werkgeheugen wordt geconfronteerd met de kennis die reeds in het lange termijngeheugen aanwezig en bewerkt is. De capaciteit van het langetermijngeheugen om kennis en vaardigheden op te slaan in een hiërarchisch netwerk lijkt onbeperkt. Van Merriënboer en Sweller (2005) onderscheiden drie soorten cognitieve belasting:

- *Intrinsieke cognitieve belasting* (intrinsic cognitive load): heeft betrekking op de complexiteit van de leerstof, dat wil zeggen: het aantal informatie-elementen dat tegelijkertijd verwerkt moet worden. Deze cognitieve belasting is inherent aan het te leren materiaal en kan niet gewijzigd worden door de aard van het onderwijskundig ontwerp. Zeer complexe materie waarin vele nieuwe en moeilijke begrippen en concepten moeten worden geleerd, blijft complex, hoe ze ook wordt gepresenteerd.
- *Ineffectieve cognitieve belasting* (extraneous cognitive load): ontstaat wanneer studenten zich bezig houden met activiteiten die niet bijdragen tot leren. Het is een belasting die een gevolg is van het instructiemateriaal dat gebruikt wordt om nieuwe informatie te presenteren aan de lerende (bijvoorbeeld beginnende studenten die nog geen relevant schema bezitten te laten leren met trial-and-error probleemoplossen of het zoeken van informatie). De beschikbare capaciteit wordt dan vooral bepaald door de beperkte werkgeheugencapaciteit. Een aangepast onderwijskundig ontwerp kan deze cognitieve belasting wel reduceren en zo het leren vergemakkelijken.
- *Effectieve cognitieve belasting* (germane cognitive load): wordt veroorzaakt door cognitieve processen die direct relevant zijn voor leren, met name voor schema-automatisering en schemaconstructie.

Van Merriënboer en Sweller (2005) hanteren als basisprincipe: minimaliseer 'ineffectieve cognitieve belasting' en optimaliseer 'effectieve cognitieve belasting', binnen de grenzen van de totale cognitieve capaciteit. Dit komt erop neer de cognitieve belasting in het leerproces zo af te stemmen dat de interactie tussen werkgeheugen en langetermijngeheugen geoptimaliseerd wordt. Leren wordt gefaciliteerd door het vrijmaken van capaciteit voor effectieve cognitieve belasting, waardoor schemata tot stand komen die achteraf door oefening en herhaling geautomatiseerd worden en uiteindelijk opgeslagen in het langetermijngeheugen. Alle factoren die het directe leerproces kunnen bedreigen waardoor ineffectieve cognitieve belasting wordt opgeroepen, moeten worden vermeden. Op die manier wordt er meer ruimte gecreëerd voor de effectieve cognitieve belasting, waardoor het leerproces feitelijk plaatsvindt.

Het leerproces kan daarbij ondersteund worden door specifieke didactische methoden zoals uitgewerkte voorbeelden, het 'completion strategy principle' (de opdracht om taken te completeren), het 'self explanation principle' (een methodiek om zelf kritische

en relevante vragen te stellen), het 'modaliteitsprincipe' (gebruik maken van meerdere media, waardoor het sensorische subsysteem van het werkgeheugen wordt betrokken bij het optimaliseren van de beschikbare capaciteit) en het 'component fluency principle' (het automatiseren van steeds terugkerende vaardigheden door deeltaak-oefeningen). Voor de statistiek wordt ernaar gestreefd de complexiteit te reduceren door de inhoud van dat vak over een groter aantal cursussen te verdelen. Studenten die zich nog een mentaal schema moeten vormen van het statistisch begrippenapparaat, worden dan niet meer blootgesteld aan een veelheid van begrippen in een enkele cursussen. Veel beginnende studenten hebben moeite de vele statistische maten en technieken uit elkaar te houden die door de verschillende meetniveaus worden onderscheiden. Door in een eerste cursus de statistiek te beperken tot statistische maten en technieken van één meetniveau, bijvoorbeeld door de statistische inhoud bij het vergelijken van groepen te beperken tot vergelijken van gemiddelden (beschrijvend en inductief: t-toetsen/ANOVA). Pas later, nadat ze deze maten en technieken frequent hebben leren toepassen in betekenisvolle contexten wordt overgegaan op een volgende serie vergelijkende maten en technieken van een ander meetniveau en aangekoppelde toepassingen.

### 3.2.2 *Meaningful Learning*

Ausubel (1960, 1968) is vooral geïnteresseerd in de manier waarop lerenden betekenisvol materiaal verwerken. Een belangrijk leerproces is subsumptie: nieuwe kennis wordt gerelateerd aan kenniselementen in bestaande cognitieve structuren; de voorkennis beïnvloedt leren het meest. Ausubel maakt een onderscheid tussen enerzijds receptief en ontdekkend leren en anderzijds 'meaningful learning' en 'rote learning'. Het eerste onderscheid heeft te maken met de manier waarop de leerstof wordt aangeboden: bij receptief leren krijgt de lerende de stof gestructureerd aangeboden; in het andere geval moet hij zelf ontdekken wat de nieuwe leerstof inhoudt. Bij meaningful leren sluit de nieuwe lesinhoud nauw aan bij de aanwezige kennis, bij rote learning gaat het om uit het hoofd leren van feiten los van de bestaande kennis. Ausubel (1968) is een voorstander van 'meaningful reception learning': om daartoe te komen is het uitermate belangrijk de aanwezige voorkennis te activeren. Hij stelt het volgende: 'If I had to reduce all of educational psychology to one principle, I would say this: the most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly'. Als strategie om leren en onthouden van nieuwe informatie te vergemakkelijken, stelt Ausubel de 'advance organizer' voor. (Ausubel, 1968, p. 81). De belangrijkste functie van een advance organizer is dat het leermateriaal op voorhand gestructureerd wordt, waardoor de lerende zich beter kan oriënteren in de leerstof en gemotiveerd is om tot betekenisvol leren te komen.

De elaboratietheorie van Reigeluth (1992) bouwt voort op de advance organizers van Ausubel. Het belangrijkste principe van deze theorie is dat er in de instructie een toenemende mate van complexiteit nodig is om optimaal leren te verzekeren: kennisstructuren moeten opgebouwd worden van eenvoudig naar complex en er moet vooraf informatie gegeven worden over de lesinhoud die volgt. Daardoor ontstaan er

stabiele cognitieve structuren, zal transfer van het geleerde in nieuwe situaties vlotter optreden en de motivatie worden verhoogd. De lerende moet een betekenisvolle context creëren waarin nieuwe kennis kan worden geïntegreerd.

### 3.2.3 *Cognitive Apprenticeship Instruction Model (CAIM)*

In dit proefschrift krijgen de ideeën van het Cognitive Apprenticeship Instructiemodel (CAIM) een belangrijke rol toebedeeld. Niet in het minst omdat het CAIM zich richt op het verwerven van expertprocessen bij complexe taken; een belangrijk item in de beoogde onderzoekscompetentie van de psychologiestudent aan de Open Universiteit Nederland. De Cognitive Apprenticeship of leerling-gezel-meester-instructiebe-nadering werd ontwikkeld door Collins, Brown en Newman (1989) en hangt sterk samen met 'situated learning'. Brown, Collins en Duguid (1989) lichten het begrip 'Situated Learning' als volgt toe: kennis kan beter bekeken worden als een verzameling vaardigheden, die pas ten volle begrepen kan worden door ze te gebruiken in de daarbij passende context. Studenten verwerven maar al te vaak gedecontextuali-seerde definities en algoritmes, die ze niet kunnen toepassen en die dus uitmonden in inerte kennis.

Lave (1988) stelt dat leren een functie is van de activiteit, de context en de cultuur waarbinnen het leren plaatsvindt. Lave en Wenger (1991) beklemtonen dat sociale interactie essentieel is voor situated learning. Lerenden nemen deel aan een 'community of practice', een praktijkcontext, waarin ze starten aan de periferie van het beroepsdomein en waar ze met de hulp van collega's en docenten kunnen groeien om eventueel op termijn zelfs de rol van expert te kunnen vervullen. Dit is wat zij het proces van 'legitimate peripheral participation' noemen. Vanuit het perspectief van gesitueerd leren is het ondenkbaar de lerende, het te leren materiaal en de context waarin het leren gebeurt, te scheiden.

Cognitive apprenticeship is een instructiemodel dat tot doel heeft de processen aan te leren die experts gebruiken bij het uitvoeren van *complexe* taken. Collins et al. (1989) stellen dat '...before schools appeared, apprenticeship was the most common means of learning and was used to transmit the knowledge required for expert practice in fields from painting and sculpting to medicine and law. Even today, many complex and important skills, such as those required for language use and social interaction, are learned informally through apprenticeship-like methods--that is, methods not involving didactic teaching, but observation, coaching, and successive approximation' (ibid, p. 453).

CAIM verschilt van het traditionele leerling-gezel-meestermodel in die zin dat het de cognitieve en metacognitieve vaardigheden benadrukt. Taken en problemen moeten dus zo gekozen worden dat studenten het belang van bepaalde technieken in de be-roepspraktijk kunnen onderkennen en ze in verschillende settings kunnen toepassen. De moeilijkheidsgraad van de taken moet geleidelijk opgevoerd worden zodat de samenstellende vaardigheden goed geïntegreerd kunnen worden (Collins et al., 1989, p. 459) en studenten geleidelijk 'taakvolwassen' worden. Brown, Collins en Duguid

(1989) veronderstellen voorts dat leersituaties die een duidelijke overeenkomst met de authentieke werksituatie vertonen het meest geëigend zijn om de transfer van leer- naar werksituatie te bevorderen.

Collins, Brown en Holum (1991) vatten hun ideeën over succesvol onderwijs samen in een algemeen raamwerk voor het ontwerp van leeromgevingen. In hun Cognitive Apprenticeship Instructiemodel (CAIM) onderscheiden zij vier dimensies: inhoud, methode, opbouw van de leertaken ('sequencing') en sociale context.

#### *Inhoud*

Collins et al. (1991) benadrukken dat in de leeromgeving strategische kennis aanwezig moet zijn. Daarmee verwijzen zij naar de impliciete kennis waarover een expert beschikt om bij de uitvoering van taken gebruik te maken van concepten, feiten en procedures die nodig zijn om problemen op te lossen. Deze expertkennis bestaat uit domeinkennis, heuristische strategieën, controlestrategieën en leerstrategieën.

#### *Methode*

De auteurs stellen zes instructiestrategieën voor om het leren te bevorderen. De eerste drie (modeling, coaching en scaffolding/fading) vormen de kern van CAIM en helpen studenten om vaardigheden te verwerven via observatie en begeleide praktijk. De volgende twee methoden (articulatie en reflectie) helpen studenten om gericht te observeren hoe de expert problemen oplost en bewust hun eigen probleemoplossende strategieën te controleren. De laatste methode (exploratie) heeft als doel de zelfstandigheid van de student aan te moedigen, niet alleen in het uitvoeren van expert probleemoplossingsprocessen maar ook in het definiëren en formuleren van de problemen die moeten worden opgelost.

- *Modelling*: de expert voert de taak uit, de lerende observeert en vormt zich een conceptueel model van de processen die vereist zijn om de taak tot een goed einde te brengen. De interne processen, heuristieken en controleprocessen waarmee experts hun conceptuele en procedurele kennis toepassen, moeten daarbij veruiterlijkt worden, zowel in gedrag als in begeleidende uitleg.
- *Coaching*: de expert observeert de studenten bij de uitvoering van een taak en geeft hints, feedback en nieuwe taken die de prestaties van de student dichterbij de buurt moeten brengen van die van de expert. Inhoudelijk gezien moet de interactie rechtstreeks samenhangen met pogingen van de student om de taak tot een goed einde te brengen. Voor het docententeam om de student in zijn of haar ontwikkeling te volgen.
- *Scaffolding en fading*: de expert ondersteunt (scaffolding) de student bij de uitvoering van een taak, vereenvoudigt of past de taak aan en vermindert zijn ondersteuning (fading) naarmate de lerende zelfstandiger wordt. Scaffolding vereist een accurate diagnose van het huidige niveau van de vaardigheden, of met de huidige terminologie het competentieniveau, van de student.
- *Articulatie*: de student wordt aangemoedigd om zijn kennis, zijn manier van redeneren en probleemoplossende processen te expliciteren.

- *Reflectie*: studenten kunnen hun eigen probleemoplossende processen vergelijken met die van een expert of andere studenten.
- *Exploratie*: naarmate ze meer expertise opdoen, moeten studenten ertoe gebracht worden een eigen manier van probleemoplossing te ontwikkelen. Studenten aanzetten tot exploratie is cruciaal als ze willen leren hoe vragen te stellen en probleemsituaties aanpakken. Exploratiestrategieën zouden dan ook moeten worden aangeleerd als onderdeel van algemene leerstrategieën.

#### *Opbouw van de leertaken*

Studenten moeten taken krijgen die zo zijn opgebouwd dat hun leerproces gestructureerd wordt en tegelijkertijd betekenisvol blijft. Het is belangrijk dat de volgende drie principes daarbij in balans blijven:

- opbouw volgens stijgende complexiteit (zie ook Reigeluth, 1992)
- opbouw volgens stijgende diversiteit
- opbouw van globale naar beroepsspecifieke vaardigheden.

#### *Sociale context*

Binnen de sociale organisatie van een ‘apprenticeship’ starten lerenden meteen met activiteiten die rechtstreeks met de productie te maken hebben. In de context van realistische problemen, met de experts als voorbeeld, verwerven ze snel de nodige vaardigheden. Collins et al. (1991) onderscheiden vier belangrijke kenmerken in de sociale context van het leren.

- *Gesitueerd leren*: de omgeving waarin studenten taken uitvoeren moet een beeld geven van de verschillende manieren waarop hun kennis in de toekomst kan worden aangewend. Studenten kunnen op die manier leren wanneer ze een bepaalde strategie al dan niet kunnen gebruiken. Leren in verschillende contexten bevordert bovendien de transfer naar nieuwe problemen en domeinen.
- *Community of practice*: de leeromgeving moet de student de mogelijkheid geven om met medestudenten en een expert actief te communiceren over en bezig te zijn met de praktijk van probleemoplossing in een bepaald domein.
- *Intrinsieke motivatie*: binnen de leeromgeving moet intrinsieke motivatie gepromoot worden door interessante en authentieke taken.
- *Samenwerking*: de leeromgeving moet studenten mogelijkheden bieden om samen opdrachten en problemen aan te pakken.

#### 3.2.4 *Enkele empirische bevindingen van CAIM*

In deze paragraaf worden de resultaten van een aantal onderzoeken beschreven, waarin CAIM is toegepast. Daarbij wordt vooral gericht op de onderzoeksresultaten en de mate waarin deze evidentie opleveren voor de veronderstelde effectiviteit van het CAIM.

Willemsen en Gainen (1995) introduceerden het CAIM in een elementaire statistiekcursus op bachelorniveau. De studenten kregen zorgvuldig ontwikkelde real-live cases en werden daarbij ondersteund door de docenten (scaffolding). In een gesimuleerde

expertpraktijk en met rollenspelen verwierven de studenten meer verfijnde opvattingen over het gebruik van statistiek, pasten ze vlotter statistische interpretaties bij een gevarieerd publiek aan, en verwierven zij de nodige vaardigheden om kritische consumenten te zijn van onderzoeks- en mediaboodschappen.

Om bachelorstudenten onderwijskunde wetenschappelijke en statistische 'literacy' bij te brengen ontwierpen Derry, Levin, Osana, Jones en Peterson (2000) een innoverende cursus statistiek volgens de principes van gesitueerd leren en cognitive apprenticeship. Van de studenten werd verwacht dat ze in kleine groepjes slecht gestructureerde (ill-structured) en authentieke problemen aanpakten om vervolgens de resultaten in een klasconferentie te presenteren en te verdedigen. In die kleine groepjes werden onderzoeksprocessen aangemoedigd en zorgvuldig begeleid door docenten en assistenten (scaffolding). De cursus was op veel vlakken een succes. Kwalitatieve en kwantitatieve analyses toonden aan dat de studenten door de cursus beter in staat waren onderzoeksrapporten en statistieken in de media kritisch te bekijken en te commentariëren. Anderzijds bleek uit de kwalitatieve analyses ook dat er een bepaalde conceptuele verwarring optrad, volgens de auteurs te wijten aan het ontwerp van de cursus. Opvallend waren ook de uiteenlopende affectieve reacties; in het algemeen steeg de tevredenheid over de cursus in de loop van twee semesters. Sommige studenten hadden het niet begrepen op de samenwerking met andere studenten; anderen klaagden over de enorme hoeveelheid werk die de cursus met zich meebracht. Ook de docenten brachten naar voren dat het begeleiden van studenten in een authentieke probleemoplossende context een veeleisende, kostbare en tijdrovende opdracht is in vergelijking met het traditionele onderwijs.

Het onderzoek van Hendricks (2001) richt zich op de vraag of gesitueerd leren meer transfer van kennis oplevert dan instructie die losstaat van de context waarin die kennis gebruikt moet worden. Om die vraag te beantwoorden kregen 220 grade 7-studenten instructie over causaliteit; ze leerden hoe te bepalen of een onderzoeksstudie al dan niet een oorzaak-gevolg-relatie vertoonde. Een helft van de studenten werd geïnstrueerd in een CAIM-conditie; de andere helft in een klassieke lessituatie. De studie toonde aan dat in de groep die instructie had gekregen in een CAIM-conditie tweemaal zoveel studenten een goed begrip hadden van het causaliteitsconcept in vergelijking tot de klassieke lessituatie. Hoewel er duidelijk betere onmiddellijke leereffecten waren, kon de studie niet bewijzen dat gesitueerd leren resulteert in een betere transfer naar real-lifesituaties. Een opvallende bevinding daarnaast was dat studenten onder de CAIM-conditie actiever en gemotiveerder waren dan de studenten in de andere groep.

De Jager, Reezigt en Creemers (2002) onderzochten de effecten van een lerarentraining gebaseerd op de nieuwe inzichten van CAIM. Ze vergeleken de prestaties van twee groepen studenten, waarvan de ene getraind werd in een CAIM-conditie. De studenten uit de CAIM-omgeving veranderden op een succesvolle manier hun instructiegedrag en vertoonden een duidelijke kwalitatieve verbetering in hun instructievaardigheden.



Osana en Seymour (2004) implementeerden een CAIM-omgeving in een klas leraren om hun argumentatievaardigheden en kritisch denken over complexe onderwijskundige problemen te verbeteren. De cursus ter voorbereiding van het leraarschap duurde vijf weken. Uit kwalitatieve data-analyse bleek dat studenten die deelnamen aan de interventie beter in staat waren onderzoeksbevindingen te gebruiken bij het nemen van beslissingen over complexe problemen gerelateerd aan de schoolgemeenschap. Verder was er een bescheiden verbetering merkbaar wanneer studenten een onderscheid moesten maken tussen kwaliteit en type van onderzoeksbewijzen. Ten slotte ontwikkelden ze meer uitgewerkte ideeën over onderzoeksbewijzen en waren ze beter in staat het concept correlatie te gebruiken in hun eigen bewijsvoering.

Darabi (2005) creëerde een authentieke leeromgeving gebaseerd op CAIM-principes voor een beperkte groep masterstudenten systeemanalyse. De studenten namen de rol aan van consultants en werkten met cases, aangeleverd door echte klanten. Het plannen en implementeren van deze toepassing vereiste extra tijd en inspanning van de docenten, vooral de contacten met de klanten. De studenten vonden de cursus intellectueel gezien een veel grotere uitdaging dan studenten buiten een CAIM-omgeving en ze gaven aan dat de cursus een belangrijke bijdrage had geleverd aan hun persoonlijke leerdoelstellingen en de voorbereiding op hun toekomstige loopbaan. De studenten in de CAIM-omgeving presteerden beter dan studiegenoten in een vorig semester; de prestaties van de studenten werden bovendien bevestigd door de uiterst positieve feedback van de klanten.

Bovengenoemde bevindingen lijken veelbelovend en geven aanleiding om in het ontwerp voor de onderzoekscompetentie ruime aandacht aan instructieaanwijzingen uit CAIM te besteden.

### 3.3 **Theory of the Research Training Environment (Gelso, 2006)**

In de hoofdstukken 1 en 2 is al opgemerkt dat een negatieve attitude en een gebrek aan motivatie schering en inslag zijn wanneer studenten geconfronteerd worden met statistiek. In zijn Theory of the Research Training Environment (RTE) stelt Gelso (2006) tien 'research-enhancing (and retarding) ingredients' voor om de attitude en motivatie van psychologiestudenten ten opzichte van onderzoek en daarmee van statistiek te verbeteren (p. 6). De eerste zes daarvan worden volledig ondersteund door wetenschappelijk onderzoek, de laatste vier onder bepaalde voorwaarden.

De eerste component die Gelso noemt is een cultureel kenmerk waarin studenten worden ondergedompeld. Hij stelt dat de opleiding zelf het voorbeeld voor wetenschappelijk gedrag en attitude moet geven. Docenten moeten aan studenten tonen dat onderzoek doen plezier en voldoening verschaft. Als effectief rolmodel moet de opleiding niet alleen de successen maar ook de mislukkingen delen met de studenten. Zo krijgen studenten de mogelijkheid om te ervaren dat ook de besten kunnen falen en krijgen zij meer zicht op de haalbaarheid van de eigen onderzoeksdoelen.

Als tweede component adviseert Gelso dat de opleiding studenten geïnteresseerd en enthousiast dient te maken voor onderzoek – zowel formeel als informeel – en de onderzoeksinspanningen van studenten positief te bevestigen. Die bevestiging kan zeer concreet zijn in de vorm van prijzen voor studentenonderzoek of mogelijkheden om conferenties bij te wonen. Wat echter het meest effect blijkt te scoren, is de interpersoonlijke erkenning, vooral in een context van tutor-tutee zoals dat ook in het CAIM aan bod komt. Het belang van een dergelijke relatie is aangetoond in verscheidene studies.

Een derde component is om de student zo vroeg mogelijk in de opleiding en op een zo veilig mogelijke manier bij onderzoek te betrekken. De eerste kennismaking met onderzoek is vaak een cursus statistiek. Zo ontstaat volgens Gelso – zeker voor niet-wiskundig georiënteerde studenten – het idee dat onderzoek statistiek is, een harde noot om te kraken voor velen (zie tevens Mellenbergh, 1980; Verschuren, 2004). Statistiek die niet kadert in een onderzoekscontext noemt Gelso weinig betekenisvol. Om de onderzoekscompetentie naar een hoger niveau te brengen, is het belangrijk dat studenten zo snel mogelijk kunnen proeven van de ‘opwindende’ aspecten van onderzoek en wetenschap. Op die manier vermindert de angst en onzekerheid van de studenten over zichzelf als onderzoeker en wetenschapper. Verschillende studies bevestigen een matig tot sterk verband tussen een vroege betrokkenheid in het onderzoeksgebeuren tijdens hun studie en een positieve attitude tegenover onderzoek in hun latere carrière.

Als vierde stelt Gelso dat tijdens de onderzoekstraining benadrukt moet worden dat alle onderzoeksstudies hun beperkingen hebben. De belangrijkste oorzaak voor de geringe interesse van studenten voor onderzoek is hun overtuiging dat ze feilloos onderzoek moeten verrichten met wereldschokkende resultaten. De RTE moet de student duidelijk maken dat elk onderzoek zijn beperkingen heeft, maar dat onderzoek hoe dan ook wel bijdraagt aan kennisvermeerdering. Die wetenschap zal voor de student een hele opluchting zijn en een opsteker voor hun motivatie. Diezelfde opvatting brengen Collins et al. (1991) naar voren wanneer ze het belang van modeling onderstrepen: ‘...Even experts stumble, flounder, and abandon their search for a solution until another time. Witnessing these struggles helps students realize that thrashing is neither unique to them nor a sign of incompetence’.

Om de attitude en een efficiënte onderzoeksaanpak van studenten positief te beïnvloeden, moeten ze volgens Gelso ten vijfde vertrouwd worden gemaakt met diverse – zowel kwantitatieve als kwalitatieve – onderzoeksmethoden. Daardoor kunnen studenten zich vrijer voelen om de bestpassende methode voor hun onderzoeksvragen te gebruiken. Gelso noemt een aantal studies waarin de positieve effecten van het aanleren van verschillende onderzoeksmethoden zijn aangetoond.

Als zesde en laatste evident gebleken component van zijn theorie noemt Gelso het advies beginnende studenten te laten ervaren hoe onderzoek en praktijk kunnen



worden geïntegreerd. Wat gecommuniceerd moet worden is dat de praktijk een belangrijke leverancier is van ideeën voor empirisch onderzoek en dat onderzoek de praktijk kan verbeteren, zij het op een indirecte manier. De indirecte relevantie blijkt een één van de nuttigste concepten om studenten te tonen hoe onderzoek de praktijk beïnvloedt. Van alle besproken ingrediënten is het empirisch bewijs voor het effect van deze laatste het grootst.

Minder evident is de component dat studenten geholpen moeten worden om zichzelf te zien als een integraal deel van het onderzoeksproces. Gelso ziet onderzoek als een proces in drie stappen: 1) de onderzoeker kijkt naar 'buiten' om voldoende kennis op een bepaald terrein op te doen; 2) de onderzoeker experimenteert door naar 'binnen' te kijken en met zijn eigen ideeën het centrum te worden van het kennisgenererende proces; 3) de onderzoeker test zijn ideeën door een aanvaardbaar onderzoeksdesign op te zetten. Gelso noemt een studie van Royalty et al. (1986) die bij 358 studenten de impact onderzochten van de component 'naar binnen kijken' en geen significant effect vonden op de onderzoeksattitude. Wanneer de populatie echter werd opgedeeld in de verschillende studiejaar kwamen ze tot de interessante bevinding dat er significante positieve correlaties werden gevonden voor de studenten van het tweede en derde studiejaar.

Ook is niet evident gebleken dat het belangrijk zou zijn dat de leeromgeving rekening houdt met de sociale en interpersoonlijke oriëntatie van studenten door enerzijds de trainer-trainee-relatie te benadrukken en anderzijds mogelijkheden te bieden om onderzoek in teamverband te doen. De docent/trainer is vooral effectief als hij de student empatisch, positief en enthousiast benadert en een evenwicht vindt tussen de rol van bedachtzame criticus en consultant-collega. Door onderzoek in teamverband te doen kan de student de positieve effecten ervaren van de interpersoonlijke relaties en meer ervaring opdoen in het doen van onderzoek, op voorwaarde dat het team geleid wordt door onderzoekers met ervaring in de vaak delicate dynamiek binnen een team (zie ook Collins et al., 1991). Onderzoek bracht aan het licht dat vooral studenten van het tweede en derde studiejaar baat hebben bij deze benadering van wetenschap als een sociaal-interpersoonlijke ervaring.

Een van de tien componenten in de oorspronkelijke theorie van Gelso (1979) was dat de artificiële band tussen statistiek en onderzoek zou worden opgeheven. Studenten psychologie zijn immers minder wiskundig gericht en de idee 'master' in de statistiek te moeten worden om te kunnen onderzoeken schrikt hen af. Een andere reden voor dat voorstel was dat het statistiekonderwijs helemaal niet inspeelde op de behoeften van de psychologiestudenten. Na 15 jaar studie van de onderzoeksleeromgeving en verschillende onderzoeken die deze component niet empirisch staven, verfijnt Gelso (2006) deze component van zijn theorie. Docenten moeten studenten duidelijk maken dat ze geen expert in de statistiek hoeven te zijn om efficiënte onderzoeker te worden. Bovendien moeten onderzoeksleeromgevingen zo zijn ingericht dat de instructie van de statistiek overeenkomt met het niveau en de behoeften van toekomstige toegepaste wetenschappers.

Als laatste component van zijn theorie stelt Gelso dat studenten vooral in het laatste deel van hun studie zouden moeten kunnen meewerken in onderzoeksbureaus en researchteams en deelnemen aan discussies in seminars. Studenten leren dan formeel en informeel het beleid, de concrete onderzoeksissues, de tijdsbewaking en de sfeer van onderzoeksbureaus. Gelso heeft geen onderzoek kunnen traceren waarin deze component is onderzocht.

### 3.4 Matrix van instructieaanwijzingen

Uit bovengenoemde cognitivistische en constructivistische stromingen zijn bijzonder veel specifieke aanwijzingen voor instructie te destilleren. Selectie van instructieaanwijzingen geschiedt op basis van de in hoofdstuk 2 genoemde veranderdoelen. Vervolgens dienen de geselecteerde instructieaanwijzingen volgens Brophy (1999) simultaan in een ontwerp te worden toegepast opdat optimaal profijt kan worden getrokken van de elkaar wederzijds versterkende effecten. Op dit punt verlaat dit proefschrift deze werkwijze van Intervention Mapping. Ten eerste is niet duidelijk in welke mate instructieaanwijzingen evidence-based zijn, een belangrijke voorwaarde voor Intervention Mapping. Ten tweede valt een gedetailleerde beschrijving van alle specifieke instructieaanwijzingen buiten het bereik van dit proefschrift. De weergave in dit proefschrift van Intervention Mapping betreft een reconstructie en niet de methodiek die is gevolgd tijdens het herontwerpen zelf. Wellicht dat nader onderzoek uit zal wijzen in hoeverre een gedetailleerde uitwerking tot een effectiever ontwerp leidt.

In bijlage I.2 zijn de (globale) instructieaanwijzingen samengevat. In het volgende hoofdstuk worden deze instructieaanwijzingen gesynthetiseerd tot een onderwijsontwerp. Terugblikkend op de geselecteerde instructieaanwijzingen, vertrekkend vanuit Gelso's Theory of the Research Training Environment tot aan de cognitivistische en constructivistische principes, kan worden vastgesteld dat het betekenisvol leren van statistiek voor psychologiestudenten impliceert dat zij vanaf het begin van de opleiding actief worden betrokken bij het uitvoeren van psychologisch onderzoek. Van meet af aan en stap voor stap, volgens CAIM en Gelso's RTE, worden studenten vertrouwd gemaakt met de rol en taken van de psychologisch onderzoeker in een variatie aan contexten, waarin verschillende onderzoeksmethoden en bijpassende statistische technieken frequent worden geoefend en toegepast en expertvoorbeelden worden gedemonstreerd en geobserveerd totdat de student zelf in staat is vorm te geven aan de opzet en uitvoering van zijn of haar onderzoek. De statistische inhoud wordt evenwichtig verdeeld over de variatie aan onderzoekstaken, waardoor niet alleen veelvuldig wordt geoefend met het toepassen van passende methoden en statistische technieken maar er tevens zorg voor wordt gedragen dat de effectieve cognitieve belasting zo optimaal mogelijk verloopt.



## **Van instructieaanwijzingen naar een geïntegreerd competentiegericht onderwijsontwerp**

### **4.1 Inleiding**

Hoofdstuk 3 is afgesloten met een uitgebreide lijst van instructieaanwijzingen die volgens de onderwijskundige literatuur effectief zijn om meer leerdoelen te realiseren. In dit hoofdstuk wordt de vraag beantwoord hoe die reeks losse instructieaanwijzingen op samenhangende wijze in een competentiegericht ontwerp kan worden ondergebracht. Door het zoveel mogelijk simultaan toepassen van instructieaanwijzingen die elkaar wederzijds versterken, wordt volgens Brophy (1999) immers de kracht achter effectief onderwijs vergroot.

In paragraaf 4.2 wordt een ontwerpmodel beschreven dat beloftevol lijkt om een krachtige leeromgeving te creëren: het Four-Component Instructional Design Model (4C/ID-model) van Van Merriënboer (1997). In dit model krijgen de diverse instructieaanwijzingen een logische samenhang.

Nadat het raamwerk voor het ontwerp van het nieuwe ‘vak’ de onderzoekscompetentie is geschetst, moet het ontwerp inhoudelijk ‘gevuld’ worden. Hiertoe worden de inhouden van methoden en technieken van onderzoek (M&T) en statistiek over de verschillende competentieniveaus verdeeld. Het resultaat van de verdeling van de leerinhouden – de onderwijsblauwdruk voor de onderzoekscompetentie – komt in paragraaf 4.3 aan bod. Ook wordt aangegeven hoe het nieuw gecreëerde vak binnen de randvoorwaarden van het modulair opgezette curriculum van de opleiding Psychologie van de Open Universiteit Nederland kan worden ingepast.

Hiermee is het ontwerp voor de onderzoekscompetentie in zoverre gereed dat deze na uitwerking daarvan in concrete studietaken en bronnen bij wijze van proef aan studenten kan worden aangeboden. Paragraaf 4.4 beschrijft de ervaringen die met het ontwerp in pilots zijn opgedaan. Deze ervaringen zijn dermate veelbelovend dat besloten is het competentiegerichte ontwerp daadwerkelijk te implementeren in het curriculum van de opleiding Psychologie. Voor een evaluatie daarvan wordt verwezen naar hoofdstukken 5 (crosssectioneel) en 6 (longitudinaal).

### **4.2 Het onderwijsontwerp**

In hoofdstuk 3 zijn onderwijskundige modellen besproken die effectieve instructieaanwijzingen hebben opgeleverd. Genoemde modellen zijn echter niet specifiek genoeg om de verzamelde instructieaanwijzingen simultaan in te bedden in één ontwerp van

een krachtige leeromgeving die bij de lerende de vereiste leerprocessen en motivatie kunnen uitlokken en op gang houden om doelmatig de beoogde leerresultaten te bereiken. Volgens Merrill (2002) is het 4C/ID-model een van de meest uitgewerkte ontwerpmodellen waarin het leren wordt gestimuleerd door oplossen van (authentieke) problemen, activering van voorkennis als basis voor verwerving van nieuwe kennis, demonstratie (modelling), toepassing en integratie van nieuwe kennis in de leefwereld van de lerende. In dit promotieonderzoek is het Four-Component Instructional Design Model (4C/ID) van Van Merriënboer gekozen als ontwerpmodel. In deze paragraaf worden eerst de vier componenten van dit model beschreven (paragraaf 4.2.1). Daarna wordt aangegeven waarom dit model geschikt is en past binnen de doelstellingen van deze interventiestudie (paragraaf 4.2.2).

#### 4.2.1 *Four-Component Instructional Design Model (4C/ID)*

Het 4C/ID-model onderscheidt zoals de naam al aangeeft, vier componenten (4C) die samen de onderwijsblauwdruk vormen: leertaken, ondersteunende informatie, 'just-in-time'-informatie en deeltaakoefeningen.

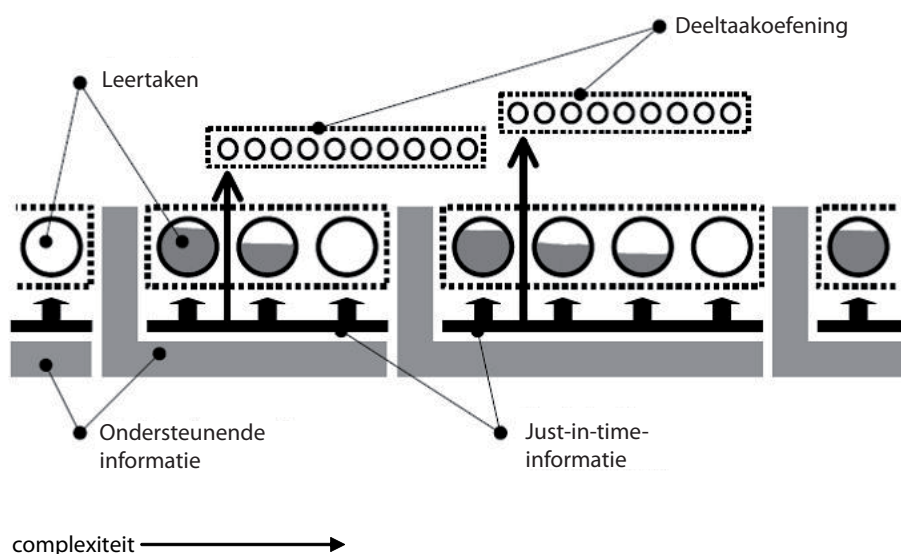
##### *Leertaken*

De belangrijkste component in het ontwerp wordt gevormd door de authentieke leertaken (zie de cirkels binnen de gestippelde rechthoeken in figuur 4.1). Deze zijn bij voorkeur gebaseerd op problemen uit de beroepspraktijk. Elke leertaak omvat de 'hele' beroepstaak en wordt in een realistische beroepssituatie uitgevoerd. Voor de onderzoekscompetentie hebben de leertaken de vorm van een onderzoekstaak.

Leertaken worden verdeeld over taakklassen (in figuur 4.1 zijn dat de gestippelde rechthoeken met daarin de leertaken als cirkels) die zijn geordend van eenvoudig naar complex. Voor de beoefening van de statistiek zijn de taakklassen te ordenen naar de mate waarin concepten en technieken op een lager niveau voorwaardelijk zijn voor concepten en technieken op een hoger niveau, bijvoorbeeld de hiërarchisch opklimmende reeks gemiddelde → standaardafwijking (variantie) → t-toets → ANOVA → ANCOVA → MAN(C)OVA → Structural Equational Modeling. Maar aangezien leertaken betrekking hebben op 'hele' onderzoekstaken waarin studenten moeten leren de nauw in relatie tot elkaar staande samenstellende elementen van die taak te coördineren en te integreren, moet deze sequentiëring eveneens gepaard gaan met toename in complexiteit van M&T- en psychologie-inhouden.

Leertaken binnen eenzelfde taakklasse die qua structuur en complexiteit equivalent aan elkaar zijn, starten met veel ingebouwde ondersteuning (de 'vulling' van de leertaken) conform de instructieaanwijzingen 'modelling', 'scaffolding' en 'fading' van het cognitive apprenticeship-model zoals beschreven in hoofdstuk 3.

Bij de laatste leertaak binnen een taakklasse vindt er geen of nauwelijks ondersteuning meer plaats. In figuur 4.1 worden de taakklassen van elkaar gescheiden door verticale, grijs getinte kolommen. De afnemende 'grijsvulling' binnen de cirkels duidt op de afnemende ondersteuningsintensiteit 'fading'. De laatste leertaak uit een taakklasse die geheel zelfstandig moet worden uitgevoerd, kan als toets worden ingezet (Hoogveld, 2003; Janssen-Noordman & Van Merriënboer, 2002; Van Merriënboer 1997).



FIGUUR 4.1 De vier componenten van het 4C/ID-model Aanpassen

#### *Ondersteunende informatie*

De benodigde declaratieve, inhoudelijk-procedurele en conditionele kennis, zoals concepten die nodig zijn om de taken te kunnen verrichten, worden aangereikt via de tweede component, de ondersteunende informatie. Vertaald in termen van psychologisch onderzoek zijn dat de psychologische, methodologische en statistische concepten en (deels ook de) methoden. De ondersteunende informatie is beschikbaar gedurende de hele taak. De informatie die voorafgaat aan de uitvoering van de taak wordt in de figuur aangeduid met de grijze kolom tussen de taakklassen en de informatie tijdens het uitvoeren van de taak met de grijze liggende staaf onder de cirkels.

#### *Just-in-time-informatie*

Voor de meer routinematige, voornamelijk technisch-procedurele kennis wordt per taak de derde component, de zogeheten just-in-time-informatie aangeboden. Bij de onderzoekscompetentie zijn dat de bronnen die technieken betreffen die steeds terugkeren bij de uitvoering van een taak en die automatisch worden uitgevoerd, bijvoorbeeld de uitvoering van een variantieanalyse (ANOVA) met behulp van het

statistische softwareprogramma SPSS. In de figuur is dit weergegeven als een zwarte horizontale balk met opwaarts gerichte pijlen.

#### *Deeltaakoefeningen*

De vierde en laatste component wordt gevormd door de deeltaakoefeningen. Deze hebben betrekking op onderdelen van de hoofdtak en kunnen veelal in 'drill & practice'-vorm worden verstrekt om een gewenst niveau van automatisering te bereiken. De computer biedt hier een belangrijke mogelijkheid om snelle feedback te verstrekken.

#### 4.2.2 *Redenen voor de keuze van het ontwerpmodel*

Voor het leerprogramma van de onderzoekscompetentie is het 4C/ID-model om de volgende redenen interessant als ontwerpmodel.

Ten eerste gaat het 4C/ID-model uit van een hele taak-aanpak. Een hele taak-aanpak houdt in dat studenten kennis, vaardigheden en attituden dienen te integreren die noodzakelijk zijn voor een effectieve taakuitoefening en die hen de gelegenheid biedt te leren hoe zij de onlosmakelijk verbonden samenstellende vaardigheden dienen te combineren en te integreren. Het doen van psychologisch onderzoek is zo'n hele taak. Hierin hebben onderdelen uit de psychologie, M&T en statistiek op functionele wijze met elkaar te maken. De voornaamste assumptie in dit proefschrift is dat door uit te gaan van onderzoekstaken waarin op een meer vanzelfsprekende en natuurlijke wijze een beroep wordt gedaan op statistische vaardigheden, de statistiek daardoor voor studenten betekenisvoller is dan bij een conventionele vakgerichte statistiekcursus.

Ten tweede maakt het model door een systeem van cumulatief opeenvolgende taakklassen een langlopende leerlijn van eenvoudig naar complex mogelijk. Hierdoor is het niet meer nodig de voor een opleiding gewenste statistische (en M&T) inhouden in één of enkele, kortlopende compacte cursussen te behandelen. De inhouden kunnen dan systematisch in kleinere delen gedoseerd worden aangeboden in de opeenvolgende taakklassen. Deze systematische verdeling reduceert niet alleen de cognitieve belasting (Van Merriënboer & Sweller, 2005; Wouters, 2007), maar komt – door de zo ontstane mogelijkheid van het behandelen en het steeds weer toepassen van concepten en technieken – ook tegemoet aan de wijze waarop studenten zich functionele kennis eigen maken (Anderson, 2000). De cognitieve belasting wordt nog meer gereduceerd door meerdere equivalente hele-taken (lees: onderzoeken) in een taakklasse op te nemen. Hierdoor kunnen de instructieaanwijzingen uit het cognitive apprenticeship-model (in het bijzonder toepassing van modelling, scaffolding en fading) nog beter tot hun recht komen. Bovendien levert het opnemen van meerdere equivalente leertaken binnen elke taakklasse nog een voordeel op: studenten worden op deze wijze geconfronteerd met een variatie aan onderzoekssituaties. Variatie is een belangrijke voorwaarde voor het bevorderen van transfer van het geleerde naar nieuwe situaties (Van Merriënboer, 1997). In tegenstelling tot de individuele, los van elkaar staande M&T- en statistiekcursussen expliciteren de opeenvolgende taakklas-



sen de hiërarchische opbouw van het onderwijsprogramma van de onderzoekscompetentie. Aan het eind van het onderwijsprogramma mag het voor studenten geen probleem meer opleveren plaats en functie van de afzonderlijke elementen uit het onderzoeksproces aan te geven.

Een eyeopener van het 4C/ID-model is de notie van achteruit ordenen ('backward chaining') (Van Merriënboer, 1997; Janssen-Noordman & Van Merriënboer, 2002). Hiermee wordt bedoeld dat bij het uitvoeren van een complexe leertaak met diverse te onderscheiden fasen – zoals het doen van onderzoek vanaf onderzoeksvraag tot conclusie – in de allereerste taakklasse de nadruk wordt gelegd op de laatste fase van de taak. In een volgende taakklasse verschuift de aandacht naar de voorlaatste fase van de leertaak waarbij de laatste fase als 'toepassing' wordt meegenomen, enzovoorts. Wordt in deze taakklasse het bestudeerde van de laatste fase uit voorgaande taakklasse ook 'meegenomen', hetzij in de vorm van toepassing, hetzij door toepassing én nieuwe leerstof, dan spreken Janssen-Noordman en Van Merriënboer van 'achteruit ordenen met sneeuwbal effect' ('backward chaining with snowballing'). Voor de opzet van een onderwijsprogramma van de onderzoekscompetentie zijn de voordelen van deze ordening voor het ontwerp van de onderzoekscompetentie legio. Om een empirisch onderzoek te kunnen uitvoeren, is kennis van het (psychologie)domein, de M&T en statistiek noodzakelijk. De hoeveelheid domeinkennis, M&T en statistiek is echter te groot om in een eerste onderzoeksrunde – die ook nog beperkt in studiebelastingomvang moet zijn – te kunnen leren beheersen. Door op het laagste competentieniveau onderzoeksvragen te formuleren die betrekking hebben op dicht bij de beleving van studenten staande ervaringen, en niet al te zeer uit te wijden over theoretisch-psychologische modellen en M&T-aspecten kan de aandacht betrekkelijk snel naar de analyse van de vergaarde gegevens geleid worden. In de analyses leren studenten enkele cruciale statistische concepten en technieken die daarvoor nodig zijn en voeren deze uit met behulp van statistische computersoftware. Op die manier ervaren ze een betekenisvolle toepassing van de statistiek in het onderzoek en dit met minder cognitieve belasting dan bij een complete, conventionele statistiekcursus. Dit is een eerste voordeel van de notie van achteruit ordenen.

Een tweede voordeel is dat studenten al in een vroeg stadium van hun studie de 'voorafschaduw' van statistische implicaties in de voorafgaande onderzoeksfasen ervaren. In een tweede taakklasse kan dan de M&T-fase meer aandacht krijgen en dienen studenten de geleerde data-analytische technieken uit de voorgaande taakklasse probleemloos toe te passen.

Deze toepassing in nieuwe onderzoekssituaties kan beschouwd worden als een derde voordeel, zowel voor het interiorisatieproces (zie hoofdstuk 2) als ter bevordering van transfer. Voor de goede orde: dat statistiek door de achterwaartse ordening in een eerste taakklasse de nadruk krijgt, wil niet zeggen dat dit gelijkgesteld kan worden aan de traditionele kennismaking met onderzoek via de vakgerichte statistiekcursussen (Gelso, 2006). Binnen de onderzoekscompetentie is statistiek duidelijk 'facilitair' om het onderzoeksproces te kunnen afronden. Dit voordeel komt nog beter tot



zijn recht door gebruik te maken van de computer. Om variantieanalyse te kunnen uitvoeren, is het niet meer nodig de statistiek zoals die in traditionele leerboeken wordt behandeld, vanaf het eerste hoofdstuk te moeten bestuderen. De langlopende leerlijn met opeenvolgende taakklassen en de achteruitordening met sneeuwbal-effect maken concentratie op onderdelen van de statistiek mogelijk, terwijl statistische softwarepakketten het tijdrovende rekenwerk uit handen nemen. Veel onderdelen uit de statistiekleerstof kunnen nu als just-in-time-informatie worden aangeboden. Hierdoor kunnen studenten zich sneller richten op de routinematige aspecten van het doen van onderzoek zonder de leerstof van de verschillende samenstellende disciplines uitputtend te moeten bestuderen.

Ten derde is het 4C/ID-model aantrekkelijk vanwege de theoretische inbedding. Zo heeft het model veel kenmerken geadopteerd van het cognitive apprenticeship instructie-model (Van Merriënboer, 1997, p. 286). Dit model wordt in dit proefschrift beschouwd als een belangrijke leverancier van effectieve instructieaanwijzingen. Hierdoor wordt de simultane inbedding van al die aanwijzingen in één ontwerp vereenvoudigd en geconcretiseerd in een competentieniveau per taakklasse (zie hierna bij de onderwijsblauwdruk, tevens bijlage I.4). Methoden uit het cognitive apprenticeship-model zoals modelling, scaffolding en fading krijgen in taakklassen met verscheidene equivalente onderzoekstaken meer reliëf. Modelling kan worden toegepast door een eerste leer-taak in de vorm van een uitgewerkt voorbeeld met expertoplossing te gieten. Van scaffolding en fading is sprake indien van de resterende leertaken in een taakklasse steeds meer aan de student wordt overgelaten. Voorbeelden hiervan zijn leertaken in de vorm van imitatieproblemen en leertaken in de vorm van aanvulproblemen.

Ten vierde biedt het 4C/ID-model gedetailleerde richtlijnen voor het ontwerpen van onderwijs. Bijlage I.3 geeft een overzicht van de tien 4C/ID-ontwikkelstappen om tot een volledig ontwerp te komen. In dit overzicht is per component een ontwikkelstap aangegeven waar een instructieaanwijzing moet worden ondergebracht. Alle instructieaanwijzingen uit hoofdstuk 3 worden met dit model gedekt.

Naast bovengenoemde voordelen blijkt het 4C/ID-model ook te passen binnen de randvoorwaarden die de Open Universiteit Nederland stelt. Deze randvoorwaarden hebben betrekking op de elektronische leeromgeving (ELO) die bij deze instelling als een conditio sine qua non voor een bijdrage aan de innovatie van het onderwijs wordt beschouwd en het modulair opgebouwde onderwijssysteem. Voor een instelling als de Open Universiteit Nederland met afstandsonderwijs zonder vaste campus met studenten 'all over the world' is ICT met zijn mogelijkheden op het gebied van uitlevering van leermaterialen, samenwerken en communiceren ongeacht plaats, tijd en tempo zeer van nut voor het creëren van een 'krachtige onderwijsleeromgeving'. Het 4C/ID-model met zijn taakklassen maakt het mogelijk de componenten zodanig te organiseren, dat ze moeiteloos kunnen worden opgenomen binnen een modulair opgebouwd curriculum dat elektronisch wordt uitgeleverd.

Een nadeel van het 4C/ID-model is dat het tot op heden weinig is toegepast voor daadwerkelijke vernieuwing van onderwijs. De aandacht is tot nog toe vooral uitgegaan naar onderzoek betreffende de effectiviteit van bepaalde componenten van het model in domeinen met goedgestructureerde problemen zoals computerprogrammering en statistische analyse (Sarfo, 2005; Van Merriënboer & Dijkstra, 1997). De uitkomsten van de diverse onderzoeken zijn echter van dien aard, dat het 4C/ID-model ook voor onderhavige studie beloftevol is wat betreft het creëren van een krachtige leeromgeving voor onderwijs voor de verwerving van de onderzoekscompetentie.

### 4.3 Naar een blauwdruk voor het M&T- en statistiekonderwijs

Nu met de keuze van het 4C/ID-model het format van het ontwerp vastligt, te weten: een serie opeenvolgende taakklassen met equivalente leertaken en 'achteruitordening met sneeuwbal effect' tussen de taakklassen, dient de leerstof van de vakken M&T en statistiek over steeds complexer wordende taakklassen verdeeld te worden. Deze paragraaf gaat nader in op de wijze waarop de hele taak van het onderzoeksproces in met elkaar samenhangende deelprocessen wordt onderverdeeld (paragraaf 4.3.1). Daarna wordt aangegeven welke onderdelen uit de M&T en statistiek deel uit moeten maken van het onderwijsprogramma voor de onderzoekscompetentie (paragraaf 4.3.2). Op basis hiervan wordt een onderwijsblauwdruk voorgesteld die aan alle eerder genoemde overwegingen en randvoorwaarden voldoet (paragraaf 4.3.3). Paragraaf 4.3.4 behandelt tot slot enkele aanvullende, niet-inhoudelijke randvoorwaarden die de inpassing van het nieuwe onderwijsprogramma binnen het bestaande psychologiecurriculum mogelijk maken.

#### 4.3.1 Onderzoekstaak als leertaak

De leertaak bij de onderzoekscompetentie is een opdracht tot het beantwoorden van een onderzoeksvraag door middel van literatuurstudie en empirisch onderzoek. In deze taak vormen onderdelen uit de psychologie, methodologie (M&T) en statistiek de samenstellende en naadloos op elkaar aansluitende componenten (Mellenbergh, 1980; Schuyten, 2001). Deze componenten komen overeen met de indeling in secties of fasen die door de American Psychological Association (APA, *Publication Manual*, 2002) voor het format voor de artikelen in wetenschappelijke tijdschriften wordt voorgeschreven: 'Introduction', 'Methods' en 'Results'. In de vierde sectie of fase passeren in 'Discussion' alle voorgaande componenten op reflectieve wijze de revue. In tabel 4.1 is aangegeven hoe de APA-secties of fasen zich verhouden tot het gedragsdoel 'toont beheersing van functionele kennis van de statistiek bij onderzoeksopdracht' (zie tabel 2.1 in hoofdstuk 2).

TABEL 4.1 Overzicht APA-indeling van wetenschappelijke artikelen gekoppeld aan onderzoeksgedragingen

<i>APA-indeling in secties</i>	<i>Gedrag</i>
Inleiding	Probleemstelling - formuleren adequate onderzoeksvraag (naar samenhang of verschil) - formuleren doelstelling Definieren van begrippen en relaties daartussen Opstellen conceptueel model Afleiden van hypothesen
Methoden	Valide operationalisering van onderzoekseenheden, variabelen en relaties daartussen Constructie van betrouwbare meetinstrumenten
Resultaten	Keuze juiste statistische techniek Toepassing met behulp van SPSS
Discussie	Correcte interpretatie van onderzoeksresultaten in het licht van de onderzoeksvraag Presentatie onderzoeksbevindingen volgens APA-regels

#### 4.3.2 *Omvang van de benodigde M&T- en statistiekkennis*

De bacheloropleiding psychologie van Nederlandse universiteiten telt 180 ECTS<sup>1</sup>-eenheden (*Onderwijsvisitatie Psychologie*, 2001, p. 46). De vakgerichte cursussen die onder de onderzoekscompetentie zullen worden geschaard, nemen bij de bacheloropleiding psychologie van de Open Universiteit Nederland 43 van de 180 ECTS-eenheden (inclusief de bachelorthesis) in beslag. Door al deze eenheden te herverkavelen over de langlopende leerlijn voor de onderzoekscompetentie voldoet de opleiding aan de minimumstandaard van 30 ECTS-eenheden ten behoeve van methodologie volgens de Visitatiecommissie (*Onderwijsvisitatie Psychologie*, 2001, p. 45<sup>2</sup>).

#### 4.3.3 *Onderwijsblauwdruk van het onderwijsprogramma voor de onderzoekscompetentie*

In bijlage I.4 is een matrix opgenomen waarin de herverdeling van de vakinhoudelijke aandachtsvelden over de diverse competentieniveaus (taakklassen) van de onderzoekscompetentie (tot en met bachelorniveau) in een langlopende leerlijn van eenvoudig naar complex in kaart is gebracht. Janssen-Noordman en Van Merriënboer (2002) duiden zo'n matrix aan met onderwijsblauwdruk. De kolommen 'Methoden' en 'Resultaten' tonen de inhouden van de vakken die tot voor kort los van elkaar werden gedoceerd, respectievelijk M&T en statistiek). De grijsvlakken in de tabel brengen tot uitdrukking op welke fase van het onderzoeksproces in een taakklasse de nadruk wordt gelegd. De verdeling van de grijsvlakken laat zien dat de 'achteruitordening'-procedure is gehanteerd. De 'plusjes' in de M&T- en statistiekkolom benadrukken

1 ECTS = Educational Credit Transfer System. 1 ECTS-eenheid = 28 uren studiebelasting (*Onderwijsvisitatie Psychologie*, 2001, p. 46).

2 Naast de 43 ECTS-eenheden ten behoeve van de onderzoekscommissie worden er ook nog 17 ECTS-eenheden aan de (psycho)diagnostische en 13 ECTS-eenheden aan de interventiecompetentie toegekend.

steeds de herhaalde toepassing ‘plus’ toevoeging van nieuwe functionele kennis. De onderwijsactiviteiten die bijdragen aan de academische vorming van de student zijn in een aparte, zogeheten niveauonafhankelijke blauwdruk van de onderzoekscompetentie vermeld. Aan deze activiteiten dient in elke taakklasse aandacht te worden besteed.

Bij de verdeling van de leerstof over de kolommen is de opbouw van het hiërarchische statistische raamwerk als uitgangspunt genomen. Zo wordt de statistiekstof in de eerste taakklassen beperkt tot concepten zoals maten en technieken die vanaf intervalmeetniveau zijn toegestaan. Nonparametrische concepten en methoden komen in latere taakklassen aan de orde. Door deze inperking worden studenten niet direct geconfronteerd met een grote hoeveelheid statistische concepten en technieken die op dat moment voor de studenten weinig betekenisvol zijn en daardoor ineffectieve, onnodige cognitieve belasting zouden kunnen opwekken. In de allereerste taakklasse worden studenten summier ingeleid in enkele casussen met een psychologische onderzoeksvraag en wordt kort uiteengezet op welke wijze de data zijn verzameld. Deze eerste taakklasse bevat vier leertaken waarin twee groepen moeten worden vergeleken door verschillen in gemiddelden op beschrijvende wijze te beschouwen. In de eerste leertaak wordt volgens het modelling-principe uit het cognitive apprenticeship-model uiteengezet hoe een expert de groepen vergelijkt. In de twee daarop volgende leertaken worden de aanwijzingen en de begeleiding bij dat vergelijken volgens scaffoldings- en fadingprincipes geëxtensiveerd en in de laatste leertaak moeten studenten op eigen kracht de vergelijking uitvoeren. Hiermee is een taakklasse afgesloten en kan naar een volgende, meer complexe taakklasse worden overgegaan. De complexiteit van de daaropvolgende taakklassen bestaat er bijvoorbeeld in dat de eerder gedane vergelijkingen nu op inductief statistische wijze (met kans- en toetsingstheorie toegespitst op vergelijking van gemiddelden) worden uitgevoerd, in taakklasse 2 tussen twee groepen (t-toets) en in taakklasse 3 tussen twee of meer groepen (variantieanalyse). Steeds wordt bij elke eerste leertaak binnen een taakklasse vertrokken vanuit een casus met een psychologische onderzoeksvraag die ‘gemodelleerd’ met een expertoplossing wordt aangeboden. Op de volgende leertaken binnen dezelfde klasse kunnen studenten hun verworven functionele kennis oefenen en toepassen, waarbij overeenkomstig de scaffolding-, fading- en coachingsmethoden steeds meer aan de zelfwerkzaamheid van studenten wordt overgelaten. In elke volgende taakklasse wordt de tot dan aangeboden kennis van de statistiek – en later ook van M&T – als bekend verondersteld. Dit gaat door tot aan de bachelorthesis waarin alle psychologie-, M&T- en statistieklerstof als bekend en intensief beoefend wordt verondersteld.

#### 4.3.4 *Inpassing van het ontwerp van de onderzoekscompetentie in het curriculum van de opleiding psychologie*

De onderzoekscompetentie waarin de ‘moeilijk leerbare vakken’ M&T en statistiek binnen de context van de psychologie worden onderwezen, kent een opbouw van eenvoudig naar complex in een langlopende leerlijn. Zo’n leerlijn impliceert een doorbreking

van de vakgerichte benadering van het bestaande curriculum. De bachelorstudent is vanaf het begin van zijn studie actief bezig met onderzoek. De langlopende, ononderbroken leerlijn maakt het mogelijk de benodigde M&T- en statistiekstof 'just-in-time' en 'just-enough' per onderzoekscyclus aan te bieden. Al 'onderzoekende' maken de studenten zich de M&T- en statistiekleerstof eigen die tijdens de bachelorfase van de psychologieopleidingen in Nederland minimaal vereist is.

Om de competenties in het bestaande curriculum van de Open Universiteit Nederland met zijn modulair opgebouwde logistieke infrastructuur te kunnen inpassen, moet nog voldaan worden aan de volgende randvoorwaarden:

- Volgens het onderwijs- en examenreglement (OER) van de Open Universiteit Nederland dient in de competentiegerichtte onderwijsaanpak een leertaak i.c. onderzoekstaak een omvang te hebben van één ECTS of een veelvoud daarvan. Aangezien de Open Universiteit Nederland in het modulair opgezette systeem rekent met modules die eenheden van 100 uur studiebelasting (of veelvouden daarvan) exclusief 20 uur tentamenvoorbereiding beslaan, dienen binnen een module meerdere leertaken te worden ondergebracht. Een module kan dan bestaan uit 4.3 ECTS (of een veelvoud daarvan). Het 4C/ID-model met taakklassen en meerdere equivalente leertaken laat zich hierin naadloos passen. Door een taakklasse de omvang van een module te geven, kunnen equivalente leertaken ondergebracht worden in een systeem van taakklassen die voor de onderzoekscompetentie (bijlage I.4) zullen worden aangeduid met 'onderzoekspracticum'<sup>3</sup>.
- Vanwege kostenbeheersing en efficiëntieoverwegingen mag de omvang van de (face-to-face-) begeleiding van de onderzoekscompetentie niet groter worden dan die van de som van de vroegere, afzonderlijke cursussen M&T en statistiek.
- Om de taakklassen i.c. onderzoekspractica in de voorgeschreven volgorde door studenten te laten volgen, dient de OER ook een aanpassing te ondergaan. De traditionele vrijheid om zelf te bepalen in welke volgorde cursussen van de Open Universiteit Nederland worden bestudeerd, kan niet meer worden gehandhaafd indien er sprake is van een langlopende leerlijn.
- Gelet op de innovatiemissie van de Open Universiteit Nederland, waarin toepassing van ICT een conditio sine qua non is (Open Universiteit Nederland: *Nota Instellingsplan 1998-2001: Verandering in perspectief*), moeten de onderzoekspractica worden ingebed in een elektronische leeromgeving (ELO). Dit is bij de Open Universiteit Nederland het in 1998 geïnstalleerde Studienet.

Met de integrale toepassing van bovengenoemde elementen in het 4C/ID-model en aanvullende voorwaarden die de Open Universiteit Nederland stelt, is – althans op papier – de aanzet tot een krachtige leeromgeving gegeven. De vraag is of dit

3 Omdat in de allereerste vier taakklassen van de onderwijsblauwdruk voor de onderzoekscompetentie wordt ingegaan op niet al te complexe en nog relatief gemakkelijke leerstof, kunnen die taakklassen worden ondergebracht in één onderzoekspracticum. Hiermee wordt voldaan aan de voorwaarde van 4.3 ECTS-eenheden per module.

onderwijsconcept daadwerkelijk toepasbaar én haalbaar is. Hiervoor zijn een aantal pilots uitgevoerd in het kader van het project Innovatie van Methoden en Technieken Onderwijs (IMTO-project 1998-2003: Van Buuren & Giesbertz, 1998; Giesbertz & Van Buuren, 2004). Hierop wordt in de volgende paragraaf nader ingegaan.

#### 4.4 Pilots met het competentiegerichte onderwijsontwerp

##### 4.4.1 Inleiding

In de periode 1999-2003 is de werkbaarheid van het onderwijs in de onderzoekscompetentie en de inpasbaarheid daarvan in de in 1998 ontwikkelde elektronische leeromgeving Studienet van de Open Universiteit Nederland in een aantal pilots onderzocht. Bij de aanvang van het project stonden elektronische leeromgevingen nog in de kinderschoenen. Zowel docenten als studenten hadden geen of nauwelijks ervaring met online studeren, laat staan dat ze samenwerkten in zo'n omgeving. De vraag was welke problemen men kon verwachten bij de overgang van face-to-face-studiebegeleiding naar begeleiding in virtuele groepen? Hoe zouden studenten over een dergelijke begeleiding oordelen? Wat voor problemen kon men verwachten bij de introductie van een nieuw medium? Daarbovenop speelde voor docenten de vraag hoe de competentiegerichte onderwijsopvatting gestalte moest krijgen. Van hen werd immers een verandering in onderwijsopvatting verwacht die getypeerd kan worden als een 'Gestaltpsychologische switch': niet meer overdracht en toetsing van vakgerichte inhoud staat centraal ('leren door te lezen'), maar taakgericht werken waarbij studenten zich al doende bekwamen in de leerstof en waarbij het geleerde tot uiting diende te komen in de producten die zij voortbrengen ('leren door te doen'). Hoe moeten met inachtneming van de langlopende leerlijn de diverse componenten van het 4C/ID-model uitgewerkt, ontwikkeld en op elkaar afgestemd worden?

In totaal zijn er acht pilots uitgevoerd die alle via Studienet aan studenten van de faculteit Psychologie zijn aangeboden. In alle pilots zijn studenten onderzoeksopdrachten in de vorm van hele-taken voorgelegd.

De eerste serie van zes pilots (1998-2001) betreffen elk een enkelvoudige leertaak. Deze pilots zijn de pilot Elektronische cursus statistiek (ECS) en de pilots IMTO-0 tot en met IMTO-4. Het voornaamste doel van deze pilots was ervaringen opdoen in het ontwerpen, ontwikkelen en uitleveren van leerstofmaterialen in de vorm van taken en bronnen (ondersteunende informatie, just-in-time-informatie en deeltaakoefeningen) en verkenning van de mogelijkheden en grenzen van de elektronische leeromgeving ten aanzien van de vormgeving en uitlevering van leerstofmaterialen, het inspelen op de actualiteit, de begeleiding van studenten en het samenwerkend leren in virtuele groepen. In deze pilots bestonden de taakklassen zoals gezegd uit een enkele leertaak. Om studenten door het onderzoeksproces heen te loodsen, werd er in elke leertaak veel nadruk gelegd op deeltaken die binnen een bepaalde termijn moesten worden uitgevoerd waarbij de uitwerking uitvoerig van feedback werd voorzien. De



virtuele groepen werden begeleid door degenen die de leertaken en bronnen hadden ontworpen.

De tweede serie pilots, te weten: de pilots IMTO-5 en IMTO-bètarun (2001-2003) zijn in hun geheel gemodelleerd volgens de in paragraaf 4.2 geschetste 4C/ID-benadering met taakklassen bestaande uit meerdere equivalente onderzoekstaken. In deze pilots is ervaring opgedaan met de instructieaanwijzingen uit het cognitive apprenticeship model (zie hoofdstuk 3) en met de inzet van reguliere (de voormalige 'face-to-face') studiebegeleiders van de bestaande M&T- en statistiekcursussen in de elektronische leeromgeving Studienet.

In de volgende paragraaf worden alleen opzet en ervaringen met deze laatste twee pilots beschreven. Reden om de eerste serie pilots met een enkele leertaak hier niet te bespreken, is dat de kinderziekten waaraan deze pilots het meest hebben geleden (in het bijzonder problemen met ICT) gaandeweg het IMTO-project grotendeels waren verdwenen. Verder zijn de leertaken en bronnen die in drie voorafgaande pilots zijn uitgevoerd, omgewerkt in respectievelijk uitgewerkte voorbeeld- en aanvultaken en hergebruikt in de pilot IMTO-bètarun.

#### 4.4.2 *Inhoud en doelstellingen van de pilots*

Na de eerste serie pilots zijn de pilot IMTO-5 (volgens de onderwijsblauwdruk van de onderzoekscompetentie de toekomstige taakklasse *Onderzoekspracticum inleiding psychologisch experiment*, nummer 5 in de blauwdruk, bijlage I.4) en de pilot IMTO-bètarun (*Onderzoekspracticum inleiding psychologisch survey* als nummer 2 in de blauwdruk, bijlage I.4) uitgevoerd.

De reden om met IMTO-5 te starten was de dringende aanbeveling van de Visitatiecommissie (*Onderwijsvisitatie Psychologie*, 2001) om in het curriculum meer aandacht te besteden aan experimenteel onderzoek. Met het ontwikkelen van een pilot die tevens in het reguliere curriculum als cursus kon worden opgenomen, werden twee vliegen in één klap geslagen. De leertaken in IMTO-5 bestaan uit drie hele-leertaken, die elk een experiment uit een bepaald gebied van de psychologie behandelen. Wat deze experimenten gemeen hebben, is het gebruik van variantieanalyse met voor- en nametingen om de data te analyseren.

In studietaak 1 krijgt de student een experiment over geheugenprocessen gepresenteerd zoals dat door een scriptiestudent is uitgevoerd, en waarvan de data opnieuw moeten worden geanalyseerd. In studietaak 2, met als onderwerp het nut van kleine klassen in het basisonderwijs, beoordeelt en ontwerpt de student een experiment, stelt hypothesen op, vult een gedeeltelijk ingevulde datamatrix aan en voert op die data statistische analyses uit. In studietaak 3 beoordeelt de student het ontwerp van een experiment over persoonlijkheid, stelt hypothesen op, voert het experiment ook daadwerkelijk uit en analyseert de door hem ingevoerde data. Voorwaarde voor deel-

name aan de pilot IMTO-5 was dat de deelnemers de beide servicecursussen statistiek – die na invoering van de onderzoekscompetentie zullen verdwijnen uit het curriculum – succesvol hadden afgerond. De noodzakelijke statistiekkennis om de data uit experimenteel onderzoek te kunnen analyseren, was dan in principe aanwezig.

In de pilot IMTO-bètarun staat vragenlijstonderzoek (survey) centraal. Er waren twee redenen van organisatorische aard om deze pilot – die als tweede in de blauwdruk vermeld staat – toch eerder dan de toekomstige eerste taakklasse te ontwikkelen. Ten eerste was in het reguliere curriculum de bestaande M&T-cursus *Vragenlijstconstructie en interviewen* met het verouderde softwareprogramma *Onderzoeksmanager* dringend aan revisie toe. Ten tweede konden in deze pilot de producten van drie vroegere IMTO-pilots – die alle rondom het vragenlijstonderzoek ('survey') waren georganiseerd – hergebruikt worden. In deze pilot bestaat de taakklasse uit vier hele-onderzoekstaken. Wat in deze surveys telkens terugkomt, is het gebruik van vragenlijsten en schaaltechnieken in een onderzoeksmodel met (minstens) drie variabelen. Tussen twee variabelen wordt een oorzakelijk verband verondersteld en dat verband wordt mogelijk beïnvloed door een derde, de modererende variabele.

In de eerste studietaak onderzoekt de student in hoeverre de variabele sekse een modererend effect heeft op de relatie hechting en psychische gezondheid. De tweede studietaak heeft als onderwerp de relatie tussen kenmerken van werk en arbeidstevredenheid. Uit de bijgeleverde onderzoeksdata kiest de student een enkel arbeidskenmerk dat arbeidstevredenheid beïnvloedt en onderzoekt of de relatie voor de mannelijke respondenten anders is dan voor de vrouwelijke. Daarnaast onderzoekt de student nog een tweede conceptueel model waarin, in plaats van de variabele sekse, een persoonlijkheidskenmerk als moderator fungeert. In de derde studietaak reconstrueert de student het onderzoek van een scriptiestudent. Voor de vierde studietaak dient de student zelf een onderzoeksvraag te bedenken over een causaal verband tussen twee variabelen met een moderator, in de psychologische literatuur naar een antwoord te zoeken en dit te toetsen op basis van data die verzameld zijn bij een groep personen van minimaal twintig respondenten bij wie een vragenlijst moet worden afgenomen. Dat antwoord moet uiteindelijk gegeven worden in een artikel volgens het APA-format. Dat artikel is tegelijkertijd de tentamenopdracht. Om aan deze pilot te kunnen deelnemen was de succesvolle afronding van de cursus 'beschrijvende statistiek' een vereiste.

Dat IMTO-5 een studietaak minder bevat dan IMTO-bètarun heeft te maken met het feit dat de experimentele opzet en de dataverzameling bij de derde studietaak (tegelijkertijd de eindopdracht voor het tentamen) met voor- en nametingen meer tijd in beslag nemen dan de omvang van één ECTS-eenheid. Daarnaast moet er meer ondersteunende en just-in-time-informatie (zie paragraaf 4.2) over enkele meer geavanceerde variantieanalytische technieken worden toegevoegd omdat de cursussen statistiek daarin niet voorzien.



Beide pilots startten met een gezamenlijke bijeenkomst van deelnemende studenten, begeleiders en onderzoekers in een studiecetrum van de Open Universiteit Nederland. Tijdens die bijeenkomsten werden doelen en werkwijzen uiteengezet, ondersteunende schriftelijke en elektronische informatie uitgereikt en de werking van ELO gedemonstreerd. Van alle deelnemers werden foto's gemaakt die vervolgens op de site van de pilot werden geplaatst. Ook voor de afsluiting van een pilot kwamen alle deelnemers weer in een studiecetrum bij elkaar. Deze afsluitende bijeenkomst stond in het teken van de evaluatie van de gehele pilot. Tussen deze twee face-to-face-bijeenkomsten door verliepen de contacten en uitwisselingen via de elektronische leeromgeving. Studietaken, ondersteunende en just-in-time-informatie en deelopdrachten werden ten dele schriftelijk en ten dele via dit medium verspreid.

TABEL 4.2 Overzicht van de pilots IMTO-5 en IMTO-bètarun met doelstellingen en looptijden

<i>Pilot</i>	<i>Doelstellingen</i>	<i>Onderwijsperiode (aantal onderwijsweken)</i>
IMTO-5	<p><b>Voor Open Universiteit Nederland</b></p> <ul style="list-style-type: none"> <li>- ontwerpen van taakklasse met equivalente onderzoekstaken volgens 4C/ID-model die met omvang van 4.3 ECTS-eenheden inpasbaar is in vakgericht psychologiecurriculum</li> <li>- toepassing cognitive apprenticeship aanwijzingen (o.a. modelling, coaching, scaffolding en fading)</li> </ul> <p><b>Voor studenten</b></p> <ul style="list-style-type: none"> <li>- zelfstandig uitvoeren van een psychologisch experiment met toepassing van meer uitgebreide variantieanalytische technieken met voor- en nameting</li> <li>- groepssamenwerking bij het formuleren van de onderzoekshypothesen en de dataverzameling</li> <li>- schrijven van een Nederlandstalig onderzoeksverslag volgens APA-richtlijnen</li> <li>- peer assessment gericht op bijdragen aan de discussies en de gezamenlijke taak</li> </ul>	<p>29-11-01 11-04-02</p> <p>(20 weken)</p>
IMTO-bètarun	<p><b>Voor Open Universiteit Nederland</b></p> <ul style="list-style-type: none"> <li>- ontwerpen van taakklasse met equivalente onderzoekstaken volgens 4C/ID-model die met omvang van 4.3 ECTS-eenheden inpasbaar is in vakgericht psychologiecurriculum</li> <li>- toepassing cognitive apprenticeship aanwijzingen (o.a. modelling, coaching, scaffolding en fading)</li> <li>- ervaring opdoen met variatie in tempo van studiebegeleiding in ELO: kort en langlopend traject</li> <li>- ervaring opdoen met variatie in groepsgrootte in ELO: vier en acht personen in een virtuele groep</li> <li>- ervaring opdoen met inzet van reguliere studiebegeleiding in ELO</li> </ul> <p><b>Voor studenten</b></p> <ul style="list-style-type: none"> <li>- zelfstandig uitvoeren van een vragenlijstonderzoek/survey</li> <li>- toepassing van diverse statistische technieken: regressieanalyse en correlatie, ANOVA</li> <li>- schrijven van een Nederlandstalig onderzoekspaper volgens APA-richtlijnen aan de hand van een voorbeeldartikel</li> </ul>	<p>17-01-2002 05-09-2002 resp. 17-01-2002 16-01-2003</p> <p>(resp. 24 kort traject en 48 weken lang traject)</p>

In tabel 4.2 zijn van de pilots de volgende kenmerken weergegeven:

- doelstelling van de pilot
- tijdpad van de onderwijsperiode
- aantal onderwijsweken dat begeleiding via nieuws- en discussiegroepen in de elektronische leeromgeving (ELO) is gegeven.

In het hiernavolgende wordt nagegaan in hoeverre de pilots leiden tot een haalbaar ontwerp voor competentiegericht onderzoeksonderwijs in de elektronische leeromgeving van de Open Universiteit Nederland.

#### 4.4.3 Methoden

##### *Participanten*

In tabel 4.3 zijn de volgende kenmerken van de studenten over de pilots vermeld:

- de wijze waarop studenten voor deelname aan de pilot werden geworven
- het aantal aanmeldingen
- het aantal participerende studenten dat daadwerkelijk het gehele traject heeft gevolgd
- het aantal geslaagden dat de pilot met succes heeft afgesloten
- het percentage uitvallers (= aantal toegelaten participierenden / aantal aanmeldingen)
- het aantal (absoluut en percentueel) geslaagden (= aantal geslaagden / aantal toegelaten participierenden)
- meest genoemde redenen waarom studenten hebben moeten afzien van deelname aan de pilot

TABEL 4.3 Enkele kenmerken van de participanten in de pilots

<i>Pilot</i>	<i>Wijze van werving</i>	<i>Aantal aanmeldingen</i>	<i>aantal participierenden (% uitvallers)</i>	<i>Aantal geslaagden (% geslaagden)</i>	<i>Reden uitval</i>
IMTO-5	Via oproep OUNL-blad Modulair en vorige pilots	38	26 (32%)	21 (81%)	Technische problemen met ELO; tijdgebrek
IMTO-bètarun	Via mailing naar studenten van de cursus M&T1	118	49 (59%)	49 (100%)	Technische problemen met ELO; tijdgebrek

In de pilot IMTO-5 hebben vooral studenten deelgenomen die al eerder kennis hebben gemaakt met de nieuwe onderwijsbenadering in een elektronische leeromgeving. De pilot IMTO-bètarun bestond voornamelijk uit nieuwkomers. Gesteld kan worden dat IMTO-5 deelnemers meer 'ervaren' zijn dan IMTO-bètarun-deelnemers. De laatsten hebben enkel de cursus *Statistiek 1* als voorbereiding gevolgd.

Niet alle participanten hebben ook de vragenlijst ingevuld. De bevindingen die in paragraaf 4.4.4 worden gerapporteerd zijn gebaseerd op 21 deelnemers aan de IMTO-5 pilot en op 49 deelnemers aan de pilot IMTO- bètarun.

#### *Methoden van dataverzameling*

Om een beeld te krijgen van de ervaringen met de pilots zijn de verslagen van de evaluaties van de IMTO-pilots gebruikt. Deze verslagen zijn gebaseerd op vragenlijsten die zijn ingevuld door degenen die met succes een pilot hebben afgerond. Een kopie van de vragenlijst is te vinden in bijlage II.1). De evaluatieonderzoeken zijn uitgevoerd door medewerkers van het Onderwijstechnologisch Expertisecentrum (OTEC) van de Open Universiteit Nederland.

De vragenlijst bestaat voor het grootste deel uit vijfpunts Likertschaalitems (1 = helemaal mee oneens tot en met 5 = helemaal mee eens) die peilen naar:

- motieven voor deelname aan de pilot (bijvoorbeeld: ‘Omdat ik samen met medestudenten kan studeren’)
- beleving van het samenwerkend leren (bijvoorbeeld: ‘Ik vind samenwerking noodzakelijk voor de studietaak’)
- begeleiding (bijvoorbeeld: ‘Ik vind dat de studiebegeleider voldoende feedback gaf op de inhoudelijke vooruitgang’) en over
- de uitvoering van de opdrachten (bijvoorbeeld: ‘Ik heb de volgorde van de opdrachten in de studietaak nauwkeurig gevolgd’).

Daarnaast is er met ja/nee-vragen onder andere gevraagd of men voldoende geïnformeerd was over enkele aspecten van de pilot (leerdoelen, onderwijsopzet, samenwerkend leren, gebruik software) en naar mogelijke bezwaren tegen online leren (zoals plaats- en tijdgebondenheid, kosten, taakgericht onderwijs, het zelf nemen van initiatieven, het maken van afspraken en de mate van zelfstandigheid die men toont). Studenten konden verbeterpunten aangeven over voorzieningen (wenselijk – onwenselijk). Ten slotte werden studenten uitgenodigd in hun eigen bewoordingen hun mening te geven over de gebruikte software, aangeboden teksten en opdrachten, het gehanteerde tijdschema, de kwaliteit van de begeleiding en een algemeen oordeel over de pilot.

#### *Analyseschema*

De evaluatieverslagen van de pilots bevatten slechts enkelvoudige beschrijvingen in de vorm van frequentieverdeling en enkele karakteristieken daarvan (gemiddelden, standaardafwijkingen) van de variabelen. In de hierna volgende resultatensectie zullen de beschrijvende bevindingen samengevat worden weergegeven.

#### *4.4.4 Resultaten*

In deze sectie worden de bevindingen weergegeven zoals die naar voren zijn gekomen uit de schriftelijke enquêtes aan het eind van de twee pilots.

*Motieven voor deelname aan de pilots*

Zowel voor de pilot IMTO-5 als IMTO-bètarun werd als belangrijke reden voor deelname de 'integratie van inhoud met methoden van onderzoek en statistiek' genoemd (zie tabel 4.4). Het feit dat studenten in beide pilots met een gemiddelde score van 4.35 op een schaal van 1 – 5 aangeven dat de integratie van inhoud met methode van onderzoek en statistiek hen motiveert om deel te nemen aan de pilot is een belangrijk signaal dat de behoefte aan een anders georiënteerd onderwijs ook bij de studenten sterk aanwezig is. Voor de pilot IMTO-bètarun werd 'het meedoen aan nieuwe ontwikkelingen' zelfs nog hoger (gemiddeld 4.39) gewaardeerd. Voor een groot deel van de deelnemers aan de pilot IMTO-bètarun was het een eerste kennismaking met het vernieuwde onderwijs. De hoge scores voor 'integratie van inhoud met methoden van onderzoek en statistiek' en 'het meedoen aan nieuwe ontwikkelingen' wijzen mogelijk op een onvrede met het traditionele statistiekonderwijs. Samenwerking met andere studenten – eerder genoemd als één van de leerdoelen voor de statistische aspecten van de onderzoekscompetentie – wordt in beide pilots eveneens hoog gewaardeerd (gemiddeld 3.96 voor IMTO-5 en 4.02 voor IMTO-bètarun). Het hebben van 'goede ervaringen met eerder pilots' was in beide pilots de minst belangrijke reden om deel te nemen.

TABEL 4.4 Redenen voor deelname aan pilots IMTO-5 en IMTO-bètarun

	IMTO-5		IMTO-bètarun	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
omdat ik samen met medestudenten kan studeren	3.96	1.19	4.02	0.88
omdat ik kan leren omgaan met de computer, internet, discussiegroepen en netmeeting	3.00	1,65	3.10	1.31
vanwege de integratie van inhoud met methoden van onderzoek en statistiek	4.35	1,15	4.35	0.99
omdat ik aan nieuwe ontwikkelingen wil meedoen	4.30	0,82	4.39	0.84
omdat ik goede ervaringen heb met deelname aan eerdere pilots	2.96	1,82	1.39	0.93

*Informatie over de pilots*

Meer dan de helft van de studenten gaf aan voldoende geïnformeerd te zijn over de leerdoelen en doelstellingen, de structuur en opzet, het samenwerkend leren en de technische aspecten van de pilot (zie tabel 4.5).

TABEL 4.5 Informatie over de pilots

	<i>IMTO-5</i> <i>Voldoende geïnformeerd</i>	<i>IMTO-bètarun</i> <i>Voldoende geïnformeerd</i>
Leerdoelen en doelstellingen	60%	55%
Structuur en opzet van de pilot	56%	53%
Samenwerkend leren	51%	57%
Techniek (hardware en software)	53%	61%

*Bezwaren binnen de pilots*

De pilot IMTO-5 leverde beduidend minder bezwaren op dan de pilot IMTO-bètarun (zie tabel 4.6). In de tabel valt vooral op dat 55% van de studenten die deelnamen aan de bètarun-pilot problemen heeft met de kosten die online studeren met zich meebrengt. Daarnaast vindt 53% van de deelnemers van IMTO-bètarun taakgericht werken lastig, terwijl niemand (0%) van hun collega's in de IMTO-5 daar een probleem mee heeft. Bij de 'ervaren' studenten worden 'tijdgebondenheid' (22%) en 'online werken' (17%) het meest bezwarend genoemd.

TABEL 4.6 Bezwaren bij het werken aan de opdrachten in de pilots

	<i>IMTO-5</i> <i>Bezwaarlijk</i>	<i>IMTO-bètarun</i> <i>Bezwaarlijk</i>
Tijdsgebondenheid	22%	41%
Online werken	17%	49%
Kosten apparatuur, provider, telefoon etc.	4%	55%
Taakgericht onderwijs: kennis en inzicht wordt niet meer rechtstreeks getoetst maar indirect via taken en bronnen	0%	53%
Noodzakelijke kennis en vaardigheden die met de computer te maken hebben	4%	45%

*Samenwerkend leren*

Over het algemeen waren de studenten positief over samenwerkend leren (zie tabel 4.7). De 'ervaren' studenten in de pilot IMTO-5 scoorden in het algemeen iets hoger op de items over het samenwerkend leren dan de 'onervaren' studenten in de pilot IMTO-bètarun. In beide pilots vond het merendeel van de studenten dat door samenwerking het inzicht in het onderwerp van de studietaak toenam. Positief in beide groepen was eveneens dat men van mening was dat door samenwerking het eindresultaat beter wordt dan wanneer men het individueel zou moeten doen. Overigens gaven studenten in beide pilots ook aan eerst zelf een manier te zoeken om de opdracht uit te voeren voordat ze hun groepsgenoten raadpleegden. Blijkbaar studeerden de studenten zelfstandig, dat wil zeggen: zochten eerst zelf naar oplossingen en stelden pas daarna vragen aan medestudenten. In beide groepen vond men het samenwerken stimulerend, maar tijdrovend. Met name in de pilot IMTO-5 vonden de studenten meer dan in de pilot IMTO-bètarun dat samenwerken extra tijd kost en extra werk heeft opgeleverd.

TABEL 4.7 Ervaringen met samenwerkend leren in een elektronische leer-omgeving

	IMTO-5		IMTO-bètarun	
	M	SD	M	SD
Ik vind samenwerking noodzakelijk voor de studietaak.	3.22	0.85	2.94	1.02
Ik vind dat de samenwerking in de groep slecht is verlopen.	2.30	1.18	2.60	1.28
Ik vind dat door samenwerking mijn inzicht in het onderwerp van de studietaak toeneemt.	3.78	0.90	3.38	1.02
Ik vind dat de samenwerking mij inzicht heeft gegeven in waar ik mee bezig ben.	3.35	1.15	3.25	1.06
Ik vind dat de samenwerking veel frustraties heeft opgeleverd.	2.78	1.35	2.35	1.19
Ik vind dat de samenwerking veel tijd heeft gekost.	3.65	0.93	2.75	1.06
Ik vind dat de samenwerking extra werk heeft opgeleverd.	3.57	1.04	2.85	0.99
Ik vind dat de samenwerking goed in de studietaak is verwerkt.	3.43	0.79	3.10	0.93
Ik vind dat in de studietaak ongeschikte opdrachten voor samenwerking zijn gekozen.	2.35	0.93	2.40	0.92
Ik vind dat er te veel moet worden samengewerkt in deze pilot.	2.35	0.98	2.17	0.93
Ik vind dat door de samenwerking het eindresultaat beter wordt dan wanneer ik het helemaal alleen zou doen.	3.83	1.15	3.33	1.06
Ik vind dat je via elektronisch samenwerken doeltreffender tot gemeenschappelijke keuzen kunt komen dan in face-to-face contact.	2.65	1.07	2.56	0.85
Ik vind dat je via elektronisch samenwerken sneller tot een gemeenschappelijke keuze kunt komen dan in face-to-face contact.	2.35	0.93	2.46	0.82
Ik vind dat de probleemoplossing die via elektronisch samenwerken wordt bereikt, kwalitatief goed is.	3.57	0.90	3.27	0.84
Tijdens samenwerken heb ik de neiging het voortouw te nemen.	3.09	0.67	2.85	0.92
Tijdens samenwerken aarzel ik om mijn argumenten voor het voetlicht te brengen.	2.04	0.56	2.15	0.87
Ik zoek zelf eerst naar een manier om de opdracht uit te voeren voordat ik mijn groepsgenoten raadpleeg.	4.30	0.56	4.02	0.73
Ik vind dat het elektronisch samenwerken stimulerend heeft gewerkt.	3.87	1.10	3.40	1.25
Tijdens het elektronisch discussiëren heb ik geen behoefte aan een lijst met aandachtspunten.	3.04	0.88	3.02	0.86
Tijdens het elektronisch discussiëren heb ik behoefte aan een elektronisch hulpmiddel om mijn ideeën in een schema te zetten.	2.65	0.65	2.75	0.81
Tijdens het elektronisch discussiëren heb ik geen behoefte aan informatie over oplossingen die eerder zijn bedacht voor het desbetreffende probleem.	2.57	0.90	2.33	0.63

Hoewel bij de start van de pilot IMTO-bètarun studenten in respectievelijk acht groepen van acht deelnemers, acht groepen van vier deelnemers en twee 'vrije' groepen waren ingedeeld, die óf het korte óf het langlopende traject hebben gevolgd, zijn over

deze opzet geen valide conclusies te trekken. Alle groepen kenden uitval en sommige deelnemers verzochten in een andere groep te mogen zitten. Ook de begeleiding was niet in alle groepen even adequaat (zie hierna bij begeleiding).

### *Begeleiding*

Tabel 4.8 laat zien dat de begeleiding in de pilot IMTO-5 positiever werd beoordeeld dan de begeleiding in de pilot IMTO-bètarun. Hierbij moeten twee kanttekeningen worden gemaakt.

De eerste is dat in de pilot IMTO-bètarun reguliere M&T- en statistiekstudiebegeleiders zijn ingeschakeld die zelf geen of nauwelijks onderzoek doen en op slechts één van de twee deelterreinen onderwijservaring hadden, terwijl de begeleiders in de pilot IMTO-5 actief bezig zijn met het doen van psychologisch onderzoek. Overigens hadden de begeleiders in beide pilots niet eerder met competentiegericht onderwijs gewerkt en kwamen beide groepen voor het eerst met begeleiding in een elektronische leeromgeving in aanraking. Van de begeleiders mag worden verwacht dat ze zich aan studenten presenteren als een 'scientist practitioner', die hun leer- en denkprocessen begeleidt en beïnvloedt (Gelso, 2006). Het gebrek aan onderzoekservaring van de docenten in de IMTO-bètarun verklaart wellicht waarom ze de studenten niet echt enthousiast konden maken.

Het merendeel van de studenten in de pilot IMTO-bètarun vond dat de studiebegeleider onvoldoende feedback gaf op de inhoudelijke voortgang en scoorde neutraal op de stelling dat de studiebegeleider voldoende hulp bood bij problemen in de groep. Daarnaast vonden de meeste studenten in de pilot IMTO-bètarun dat de studiebegeleider de groep te weinig motiveerde en enthousiast maakte. Geen enkele student vond dat de studiebegeleider te snel reageerde op vragen en opmerkingen van de groep. Geconstateerd is echter – en dit is de tweede kanttekening – dat in de pilot IMTO-5 de begeleiding zich niet heeft gehouden aan de begeleidingsafspraken. Om het 'experiment' met achteraf te moeilijk gebleken statistische analyses te redden, heeft de begeleiding studenten te veel en te vaak pasklare antwoorden gegeven op hun vragen over die analyses die ze in hun eindverslagen konden verwerken. De vraag is dan wat studenten zelf in deze pilot van de statistiek hebben opgestoken.

Volgens het merendeel van de studenten is er tijdens de introductiebijeenkomst voldoende uitleg gegeven over de inhoud van de studietaak, het inloggen in de studietaak, het opstarten van de benodigde computerprogramma's en wat er van de student verwacht werd. De begeleiding zal zeker een aandachtspunt zijn bij de implementatie van het nieuwe onderwijsconcept. Docenten zullen moeten leren hoe ze studenten kunnen motiveren en hun begeleiding zo kunnen doseren dat ze studenten volgens de scaffolding-, fading- en coachingsmethoden geleidelijk tot zelfwerkzaamheid brengen.

In IMTO-5 heeft het kostenplaatje van de begeleiding het budget overschreden, in die zin dat de omvang van de begeleiding beduidend groter was dan de face-to-face-begeleiding in de afzonderlijke cursussen M&T en statistiek.

TABEL 4.8 De begeleiding in de pilots MTO-5

	IMTO-5		IMTO-bètarun	
	M	SD	M	SD
Ik vind dat de studiebegeleider voldoende feedback gaf op de inhoudelijke vooruitgang.	4.43	0.66	2.88	1.30
Ik vind dat de studiebegeleider onze groep motiveerde en enthousiast maakte.	3.78	0.85	2.19	1.12
Ik vind dat de studiebegeleider onvoldoende hulp bood bij problemen in de groep.	1.78	0.95	3.02	1.16
Ik vind dat de studiebegeleider te snel reageerde op vragen en opmerkingen van de groep.	2.70	1.06	1.77	0.69
Ik vind dat er op de introductiedag voldoende uitleg is gegeven over de inhoud van de studietaak.	3.39	0.89	3.94	0.81
Ik vind dat er tijdens de introductiedag onvoldoende uitleg is gegeven over het inloggen in de studietaak.	2.52	0.99	2.13	1.10
Ik vind dat er tijdens de introductiedag voldoende uitleg is gegeven over het opstarten van de benodigde computerprogramma's	3.26	1.10	3.63	1.14
Ik vind dat er tijdens de introductiedag duidelijk is geworden wat van mij als student werd verwacht.	3.48	0.85	3.81	0.89
Ik vind dat de rol van de begeleiders duidelijk is uitgelegd tijdens de introductiedag	3.22	1.04	2.98	1.18

*Leerdoelen, opdrachten en bronnen*

In beide pilots blijken leerdoelen, opdrachten en bronnen voor de meeste studenten voldoende duidelijk (zie tabel 4.9). De studenten waren tevreden over logische opbouw van de studietaken en de aanwijzingen die ze kregen om de opdrachten goed uit te voeren. Alleen bij de pilot IMTO-bètarun was men minder positief over de koppeling van informatie en bronnen. Hier speelt wellicht de onervarenheid van deze groep deelnemers een rol: voor velen was het de eerste keer dat ze online opdrachten uitvoerden en de vertrouwdheid met de nieuwe informatietechnologie ontbrak.

*Uitvoeren van de opdrachten*

Het uitvoeren van de opdrachten is in het algemeen goed verlopen. De meeste studenten hebben de volgorde van de opdrachten in de studietaak nauwkeurig gevolgd en deze volgens de instructies uitgevoerd. De 'ervaren' studenten in de pilot IMTO-5 gaven aan hierbij meer hun eigen manier van werken gebruikt te hebben dan de 'onervaren' studenten in de pilot IMTO-bètarun. Bijna alle studenten in beide pilots vonden de studietaak geschikt om kennis te maken met respectievelijk het doen van experimenteel onderzoek en vragenlijstonderzoek. Daarnaast is in beide pilots het merendeel van de studenten het erover eens dat de studietaak niet als een gewone, schriftelijke cursus van de Open Universiteit Nederland aangeboden had kunnen worden (zie tabel 4.10).



TABEL 4.9 Meningen over leerdoelen, opdrachten en bronnen

	IMTO-5		IMTO-bètarun	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Ik vind de leerdoelen voor de studietaken duidelijk.	4.22	0.60	4.13	0.64
Ik vind de criteria waarop de laatste studietaak wordt beoordeeld duidelijk.	3.78	0.90	4.15	0.77
Bij alle opdrachten was het duidelijk wat ik moest doen.	3.57	0.90	3.79	0.90
Ik vind de volgorde van de opdrachten in de deeltaken logisch.	3.86	0.71	4.08	0.92
Ik vind de aanwijzingen bij de opdrachten duidelijk.	3.74	0.62	3.94	0.86
Ik heb de aanwijzingen bij de opdrachten goed kunnen gebruiken bij het uitvoeren van de opdrachten.	3.87	0.55	4.19	0.64
Ik vind het onduidelijk welke informatie bij welke bron te vinden is.	2.39	1.16	2.10	0.95
Ik vind dat de bronnen voldoende informatie bevatten om de opdrachten te kunnen uitvoeren.	3.32	1.21	4.06	0.93
Ik vind dat er voldoende bronnen zijn om de opdrachten te kunnen uitvoeren.	3.70	1.22	3.94	0.98

TABEL 4.10 Meningen over de uitvoering van de taak in de pilots

	IMTO-5		IMTO-bètarun	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Ik heb de volgorde van de opdrachten in de studietaak nauwkeurig gevolgd.	3.87	.69	4.15	0.97
Ik heb mijn eigen manier gebruikt om de opdrachten uit te voeren.	3.43	.90	2.65	1.02
Ik heb behoefte aan extra elektronische hulpmiddelen die me helpen om de opdrachten systematisch uit te kunnen voeren.	2.13	.55	2.15	0.92
Bij het uitvoeren van de opdrachten wil ik zelf bepalen welke hulpmiddelen ik gebruik.	3.17	.98	3.04	1.01
Ik vind deze studietaak geschikt om kennis te maken met het doen van onderzoek.	4.30	.47	4.08	0.85
De studietaak had ook schriftelijk als een gewone OUNL-cursus aangeboden kunnen worden.	2.43	1.04	2.79	1.05
Ik vind het prettig als de uitleg in de bronnen via Studienet wordt aangeboden.	3.74	.96	3.85	0.96

*Tijd besteed aan de studietaak*

Aan de studenten is gevraagd hoeveel tijd zij naar eigen schatting aan de pilot hebben besteed. De norm staat in beide pilots op 4.3 ECTS-eenheden (= 120 uur inclusief tentamenvoorbereiding). Het merendeel van de studenten in beide pilots scoort minder dan de norm. Liefst 34 van de 49 studenten (69%) besteedden minder dan 120 uur om de IMTO-bètarun tot een goed einde te brengen; 39 % (19 studenten) zelfs minder dan 100 uur. Bij IMTO-5 zeggen 13 van de 21 studenten (62%) minder dan de norm te scoren; 29% blijft onder de 100 uur. In beide groepen heeft een minderheid meer dan 200 uur

nodig gehad (12% van de studenten IMTO-5 en 10% van de studenten IMTO-bètarun. Deze objectieve cijfers geven duidelijk een ander beeld van de tijdsbesteding dan de perceptie van vooral de studenten in de pilot IMTO-5 over samenwerken liet vermoeden: zij vonden immers dat samenwerken extra tijd en extra werk opleverde.

#### *Waardering voor de pilots*

Er werd aan de studenten gevraagd hun waardering voor de pilot in een cijfer van 1 t/m 10 uit te drukken. Het gemiddelde waarderingcijfer was in beide pilots een 7. Bovendien gaven 18 van de 21 IMTO-5 studenten (86%) en 27 van de 49 IMTO-bètarun studenten (55%) aan deze manier van studeren (online/samenwerkend leren) ook in hun verder studietraject te willen toepassen. De studenten die aangaven dat niet te willen, hadden overigens de betreffende pilot met een voldoende gewaardeerd. Het ontbreken van een relatie tussen waarderingcijfer voor de pilot en het al dan niet op deze manier verder willen studeren is moeilijk te verklaren. Mogelijk spelen bezwaren als tijdgebondenheid, kosten van apparatuur en de perceptie van een tijdrovende samenwerking hierbij een rol.

#### 4.4.5 *Discussie en conclusie*

In dit hoofdstuk is beschreven hoe de instructieaanwijzingen uit hoofdstuk 3 simultaan kunnen worden toegepast met behulp van het gedachtengoed ontleend aan het Four-Component Instructional Design Model (4C/ID-model) van Van Merriënboer (1997). Studietaken en bronnen voor het ontwerp voor de onderzoekscompetentie zijn uitgewerkt en in pilots aan studenten aangeboden. De ervaringen met de pilots wijzen erop dat een geïntegreerde hele-taak-benadering voor de onderzoekscompetentie daadwerkelijk toepasbaar is.

Het feit dat studenten als belangrijkste motieven voor deelname aan de pilots 'integratie van inhoud met methoden van onderzoek en statistiek' en 'het meedoen aan nieuwe ontwikkelingen' aanhalen, kan erop wijzen dat studenten hierin een mogelijkheid zien om het gevreesde vak statistiek vanuit een andere, competentiegerichtere aanpak, te benaderen. Als de nieuwe onderwijsopzet erin kan slagen de angst voor statistiek sterk te reduceren en de motivatie en het enthousiasme van de studenten te verhogen, wordt één van de meest genoemde problemen in het statistiekonderwijs aangepakt.

Samenwerkend leren is zeker voor studenten in afstandsonderwijs een nieuw gegeven, dat in beide pilots toch hoog gewaardeerd wordt. De elektronische leeromgeving heeft groepsdiscussies en overleg met andere studenten gefaciliteerd. De deelnemers aan beide pilots gaven aan voldoende geïnformeerd en positief te zijn over de leerdoelen, doelstellingen en opzet van het ontwerp. De studenten weten behoorlijk om te gaan met de ondersteunende en just-in-time-informatie. Ze spreken bovendien hun waardering uit voor de studietaak, die hen in staat stelt een duidelijk beeld te krijgen van respectievelijk een vragenlijstonderzoek en een experiment. Studenten beginnen dus 'het grotere plaatje' te zien, wat met de geïsoleerde statistiekcursussen niet of veel minder het geval was.

Dat de pilot IMTO-5 beduidend minder bezwaren opleverde wat betreft kosten die online studeren met zich meebrengt in vergelijking met de pilot IMTO-bètarun (zie tabel 4.6) kan geweten worden aan het feit dat de pilot IMTO-5 voor het overgrote deel bestond uit studenten die eerder aan een IMTO-pilot hebben deelgenomen en dat het merendeel van de deelnemers van de pilot IMTO-bètarun voor de eerste keer aan zo'n pilot deelnam en nog behoorlijk moest investeren in apparatuur en internettoegang. Bezwaren doen zich nog wel voor bij studenten die voor het eerst kennismaken met onderwijs in de vorm van studietaken. Dat 53% van de deelnemers van de pilot IMTO-bètarun het werken met studietaken lastig vond terwijl niemand (0%) van hun collega's in de pilot IMTO-5 daar een probleem mee had, lijkt erop te wijzen dat bij een overstap van kennisgericht naar taakgericht leren eerst geïnvesteerd moet worden in een andere manier van studeren en werken. Het feit dat dit bezwaar is overwonnen bij IMTO-5 is een belangrijk gegeven voor de haalbaarheid van het nieuwe onderwijsconcept: de studietaak is immers een belangrijkste component in het nieuwe onderwijsontwerp.

De tijdsbesteding van zowel de studenten als de begeleiders, die toch een belangrijke randvoorwaarde is voor de implementatie van het nieuwe ontwerp, is in de meeste gevallen zelfs onder de norm van 120 uur gebleven. Wat de inhoudelijke begeleiding betreft, valt er nog wat te verbeteren, zowel op het vlak van de scholing van de docenten in de didactiek, de elektronische leeromgeving en het doen van onderzoek. De docent en de faculteit spelen immers een belangrijke voorbeeldrol om studenten warm te maken voor het doen van onderzoek. Dat de begeleiding in de pilot IMTO-bètarun niet in alle groepen door studenten als adequaat werd ervaren, kan op het conto geschreven worden van het feit dat in die pilot voor het eerst studiebegeleiders werden ingezet die én niet bekend waren met competentiegericht onderwijs én onbekend met het fenomeen elektronische leeromgeving.

Het procentueel aantal geslaagden voor de beide pilots is niet lager dan het slaagpercentage voor de traditionele statistiekcursussen (resp. 81% van de deelnemers voor de pilot IMTO-5 en 100% voor de pilot IMTO-bètarun tegen 83% van de deelnemers van de cursus *Statistiek 1* in het studiejaar 2002-2003). Het percentage uitval is voor de IMTO-bètarun vergelijkbaar met de cursus *Statistiek 1* (59% voor de IMTO-bètarun tegen 56% voor de cursus *Statistiek 1* in het studiejaar 2002-2003), maar is voor IMTO-5 gereduceerd tot bijna de helft (32%). Als reden voor de uitval wordt moeilijkheidsgraad alleen bij de traditionele cursussen statistiek vermeld; in de pilots worden dan weer technische problemen met ELO aangehaald. Als studenten niet meer afhaken omdat statistiek te moeilijk is, is dat een belangrijk pleidooi voor het nieuwe onderwijsconcept.

Bovenstaande ervaringen zijn gebaseerd op evaluatieonderzoeken van de pilots die zijn uitgevoerd door medewerkers van het Onderwijstechnologisch Expertisecentrum van de Open Universiteit Nederland. In deze evaluatieonderzoeken zijn alleen de ervaringen meegenomen van degenen die een pilot succesvol hebben afgesloten. Op vijf deelnemers in de pilot IMTO-5 na (19%) betreft dat alle deelnemende studenten.

Deze studenten blijken alsnog te hebben afgehaakt vanwege gezondheids- en tijdsproblemen, niet vanwege moeilijkheidsgraad. De niet-participerenden zijn degenen die zich wel voor deelname aan een pilot hadden aangemeld, maar die zich vanaf het begin niet hebben laten zien.

Bovengenoemde positieve bevindingen geven aanleiding om het ontwerp voor de onderzoekscompetentie in het curriculum van de opleiding Psychologie van de Open Universiteit Nederland in te passen. Wat niet uit de pilots kon worden opgemaakt, is of studenten specifiek meer begrip van de statistiek hebben gekregen. In het volgende hoofdstukken wordt daarop nader ingegaan. In die hoofdstukken worden de effecten van de geïmplementeerde competentiegerichte onderwijsopzet vergeleken met die van de traditionele, vakgerichte opzet.



## **Van ontwerp naar implementatie: crossectionele bevindingen**

### **5.1 Inleiding**

In hoofdstuk 4 is geconstateerd dat de interventie, de competentiegerichte leeromgeving voor de onderzoekscompetentie zoals dat gestalte heeft gekregen in de pilots IMTO-5 en IMTO-bètarun, werkbaar is gebleken. In dit hoofdstuk wordt nagegaan in hoeverre de verwachting die de competentiegerichte leeromgeving wekt, ook daadwerkelijk wordt ingelost. Deze verwachting is in hoofdstuk 2 in het determinantenmodel (zie figuur 2.1) uiteengezet. In het onderzoek waarover hier wordt gerapporteerd, worden twee groepen worden met elkaar vergeleken. De ene groep volgt onderwijs in de nieuwe competentiegerichte leeromgeving en de andere volgt de traditionele vakgerichte statistiekcursussen.

In de volgende paragrafen wordt verslag gedaan van de opzet, de uitvoering en de resultaten van het onderzoek naar de effecten van de interventie in het statistiekonderwijs van de opleiding Psychologie van de Open Universiteit Nederland.

#### **5.1.1 Onderzoeksvraag en hypothesen**

Het determinantenmodel dat in hoofdstuk 2 is besproken, maakt onderscheid in studentkenmerken en omgevingskenmerken. Laatstgenoemde kenmerken hebben betrekking op de leeromgeving van de student. In dit onderzoek worden twee leeromgevingen voor het statistiekonderwijs onderscheiden, te weten: de competentiegerichte en de vakgerichte leeromgeving. De verwachting bij de competentiegerichte leeromgeving is dat deze de student beter motiveert tot actieve en diepgaande bewerking van het leermateriaal wat leidt tot betere leeruitkomsten dan de vakgerichte leeromgeving. De vraag die in dit hoofdstuk centraal staat luidt dan ook:

*'In welke mate is de competentiegerichte leeromgeving effectiever in het realiseren van leerprocessen en leeruitkomsten dan de vakgerichte leeromgeving?'*

Voor het beantwoorden van deze onderzoeksvraag, zijn zeven hypothesen geformuleerd. De eerste zes hypothesen betreffen de hoogte van de scores op de drie groepen procesvariabelen (attitude, motivatie, leerstrategieën) en de leeruitkomsten. De laatste hypothese betreft de mediërende werking van de procesvariabelen in het model zelf.

*Hypothese 1*

De competentiegerichtte leeromgeving leidt tot een meer positieve attitude ten aanzien van de statistiek dan de vakgerichte leeromgeving.

*Hypothese 2*

De competentiegerichtte leeromgeving leidt tot sterkere motivatie dan de vakgerichte leeromgeving.

*Hypothese 3*

De competentiegerichtte leeromgeving leidt tot de inzet van meer diepgaande leerstrategieën dan de vakgerichte leeromgeving.

*Hypothese 4*

De competentiegerichtte leeromgeving leidt tot hogere leeruitkomsten op het terrein van de statistische kennis dan de vakgerichte leeromgeving.

Naast kennis van de statistiek als uitkomst van het leerproces zijn ook Autonomie en Afhankelijkheid in het doen van onderzoek als productvariabelen meegenomen. De hypothesen hieromtrent luidt:

*Hypothese 5*

De competentiegerichtte leeromgeving leidt tot hogere scores voor Autonomie in het doen van onderzoek dan de vakgerichte leeromgeving.

*Hypothese 6*

De competentiegerichtte leeromgeving leidt tot lagere scores voor Afhankelijkheid in het doen van onderzoek dan de vakgerichte leeromgeving.

*Hypothese 7*

De procesvariabelen attitude, motivatie en leerstrategieën mediëren de relatie tussen leeromgeving en leeruitkomsten.

## 5.2 **Methoden**

### 5.2.1 *Respondenten*

Studenten die zich in het studiejaar 2003-2004 inschreven voor een of beide statistiek-cursussen zijn in het najaar van 2004 benaderd om mee te doen aan het onderzoek ( $N = 1250$ ). Studenten die zich voor het studiejaar 2004-2005 inschreven voor het nieuwe *Onderzoekspracticum kwantitatieve data-analyse* zijn in het najaar van 2005 benaderd om deel te nemen aan het onderzoek ( $N = 769$ ). Bijgevolg zijn er twee groepen studenten te onderscheiden. De eerste groep volgt de traditionele, vakgerichte cursussen statistiek (VAK). De tweede groep heeft het competentiegerichtte onderwijs in het eerste onderzoekspracticum doorlopen (COMP). Van de 2019 verzonden vragenlijsten zijn



er 468 geretourneerd, een respons van 23%. Van de VAK-studenten zijn dat er 340 (27%) en van de COMP-studenten 128 (16%). De vraag of die lage respons vertekend werkt, is in een aanvullend onderzoek nader onderzocht. Dit aanvullende onderzoek werd uitgevoerd onder degenen die hun vragenlijst na twee maanden niet hadden geretourneerd. Het verslag van deze vervolgstudie is te vinden in bijlage III.1. Op basis van deze vervolgstudie kan worden geconcludeerd dat er geen aanwijzingen zijn dat er vertekening ('nonresponsbias') optreedt. In tabel 5.1 zijn van beide groepen enkele gegevens vermeld.

TABEL 5.1 Responsgegevens en enkele achtergrondkenmerken van de twee groepen studenten

	VAK 2003-2004		COMP 2004-2005		Totaal	
	aantal	%	aantal	%	aantal	%
Percentage geslaagden na één jaar	172	51	73	57	245	53
Percentage mannen	90	27	32	25	122	26
Gemiddelde leeftijd (standaardafwijking)	40.0 (9.3)		41.1 (9.3)		40.3 (9.3)	40.3 (9.3)

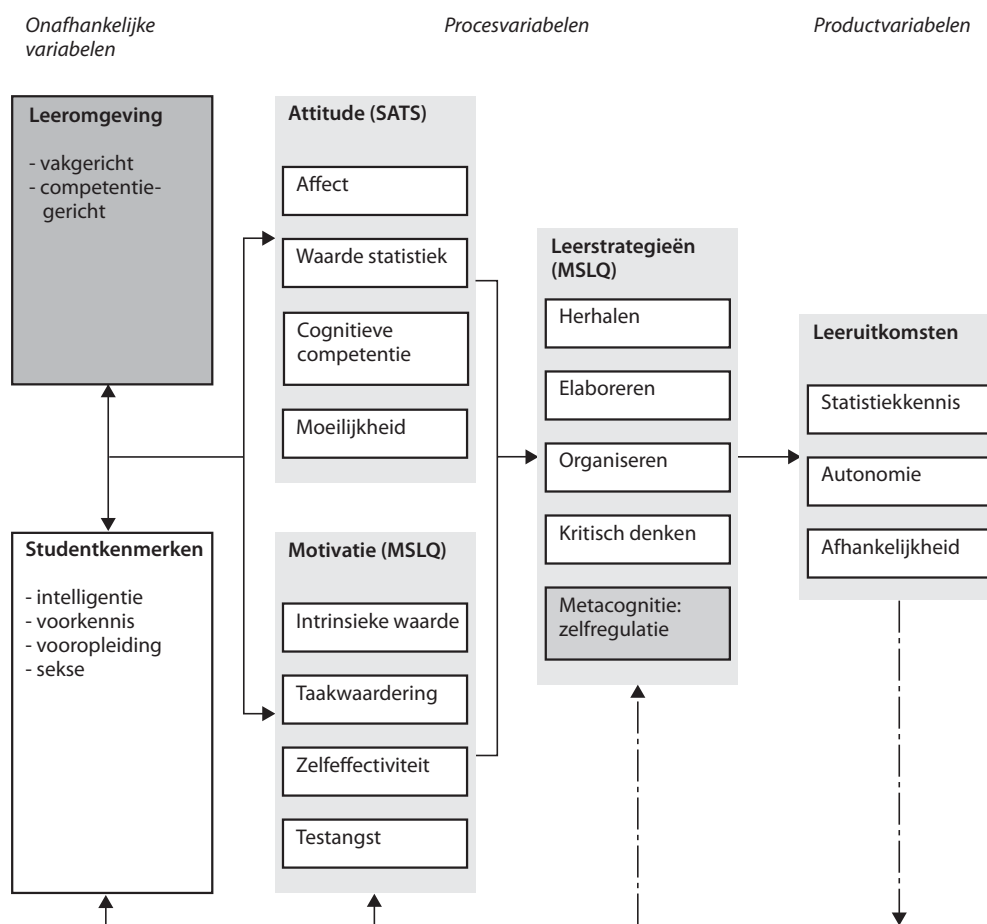
### 5.2.2 Variabelen en hun operationalisering

Voor de operationalisering van de diverse componenten in het determinantenmodel zijn overwegend bestaande meetinstrumenten (Likertschalen) gebruikt. In figuur 5.1 is het geoperationaliseerde model van figuur 2.1 met de meetinstrumenten weergegeven. Deze meetinstrumenten worden in de volgende paragrafen volgens de plaatsing in figuur 5.1 van links naar rechts beschreven.

#### *Onafhankelijke variabelen*

##### Studentkenmerken

In hoofdstuk 2 is aangegeven dat studentkenmerken als intelligentie en voorkennis – hoewel zeer relevant voor het leren – in deze interventiestudie niet zijn gemeten. Reden hiervoor zijn geweest dat (1) het niet verantwoord leek de vragenlijst nog omvangrijker te maken, (2) intelligentie onder gecontroleerde omstandigheden zou moeten worden gemeten dat gezien de aard van de studentenpopulatie van de Open Universiteit Nederland een te groot beslag zou leggen op het onderzoeksbudget, (3) het niet waarschijnlijk is dat de studenten in twee opeenvolgende jaren zeer van elkaar zouden verschillen en (4) deze variabelen als niet-veranderbare determinanten moeten worden beschouwd (hoewel ze bij bestaande groepen wel van belang zouden zijn geweest als covariaten). In figuur 5.1 zijn de relevante maar niet meegenomen variabelen met grijze letters aangegeven.



FIGUUR 5.1 Geoperationaliseerd determinantenmodel

*Kenmerken van de leeromgeving*

De in 2003 ingezette implementatie van de onderzoekscompetentielijns zal niet voor 2008-2009 voltooid zijn. Gezien dit ontwikkelproces en het gegeven dat studenten aan de Open Universiteit Nederland niet gebonden zijn aan een bacheloropleiding die in drie jaren dient te zijn afgerond, zal een evaluatie over de gehele competentielijn niet vóór 2012 kunnen plaatsvinden. In dit onderzoek worden daarom de effecten van de in 2004 uit de omloop genomen statistiekcursussen vergeleken met het in datzelfde jaar ingevoerde *Onderzoekspracticum kwantitatieve data-analyse*, het eerste deel van de onderzoekscompetentielijns. In dit practicum wordt een aantal statistische maten en toetsen behandeld dat een deelverzameling vormt van de in de vakgerichte statistiekcursussen behandelde maten en toetsen (zie tabel 5.2). De maten en toetsen in het onderzoekspracticum zijn toegespitst op het gemiddelde en de standaardafwijking, inclusief t-toetsen, enkelvoudige variantieanalyse en bivariate regressie- en

correlatieanalyse. Over de overlappende deelverzameling van technieken krijgen studenten een statistiekennistoets voorgelegd (zie hierna bij de productvariabelen). De verschillen komen in het bijzonder tot uiting in het aantal behandelde technieken, het toepassingsgericht zijn, het gebruik van statistische software (SPSS) en de toetsing in de competentiegericht benadering.

TABEL 5.2 Overeenkomsten en verschillen in de twee benaderingen van het statistiekonderwijs

Kenmerk	Vakgericht statistiekonderwijs (VAK)	Competentiegericht statistiekonderwijs (COMP)
Doelstelling	Kennis van en inzicht in van de statistiek.	Kennis van en inzicht in de toepassing van de statistiek in psychologisch onderzoek.
Instructiewijze	Studenten bestuderen een hoofdstuk uit het statistiekboek, beantwoorden vragen en maken opdrachten die toegesneden zijn op de behandelde leerstof.	In een serie equivalente onderzoekstaken (zie onderwijsblauwdruk bijlage 1.4 wordt aan de hand van vier authentieke casussen uit het domein van de psychologie volgens het 4C/ID-model (Van Merriënboer, 1997) de statistiek kort uiteengezet, beoefend en toegepast.
Aantal ECTS-eenheden	2 x 4.3 EC = 8.6 EC	4.3 EC
Inhoud	Beschrijvende en inferentiële statistiek, inclusief kanstheorie de leerstof heeft betrekking op maten en toetsen voor alle meetniveaus.	Beschrijvende en inferentiële statistiek, kanstheorie wordt minimaal en beknopt behandeld de leerstof is beperkt tot maten en toetsen voor interval- en ratiomeetniveau.
<i>Univariate statistiek</i> Frequentieverdelingen	Sectordiagram, staafdiagram, histogram, polygoon, cumulatieve verdeling, tijdreeks, boxplot.	Histogram, polygoon.
Centrummaten	Modus, mediaan, kwantielen, gemiddelde.	Gemiddelde.
Spreadingsmaten	Variatiecoëfficiënt, range, interkwartielafstand, standaardafwijking (variantie).	Standaardafwijking (variantie).
Kansverdelingen	Discrete en continue kansverdelingen: binomiaal, hypergeometrisch, Poisson, normaal, Student t, F, chikwadraat.	Continue kansverdelingen: normaal, Student t, F.
Toetsen	Alle parametrische en nonparametrische toetsen.	Alleen parametrische toetsen voor een gemiddelde (t-toetsen, enkelvoudige variantieanalyse (ANOVA))
<i>Bivariate statistiek</i> Associatiematen (en hun toetsen)	<ul style="list-style-type: none"> <li>• contingentiematen (gebaseerd op chikwadraat)</li> <li>• nonparametrische maten</li> <li>• rangcorrelatiematen,</li> <li>• correlatiecoëfficiënt <math>r</math> (<math>r^2</math>)</li> <li>• lineaire regressieanalyse</li> <li>• correlatieratio <math>\eta^2</math></li> </ul>	<ul style="list-style-type: none"> <li>• correlatiecoëfficiënt <math>r</math> (<math>r^2</math>)</li> <li>• lineaire regressieanalyse</li> <li>• correlatieratio <math>\eta^2</math></li> </ul>

<i>Kenmerk</i>	<i>Vakgericht statistiekonderwijs (VAK)</i>	<i>Competentiegericht statistiekonderwijs (COMP)</i>
Leermaterialen	<ul style="list-style-type: none"> <li>• handboeken Nijdam &amp; Van Buuren (1999)</li> <li>• OUNL-handleidingen met samenvattingen, overzichten en extra-oefenopdrachten.</li> </ul>	<ul style="list-style-type: none"> <li>• OUNL-studietakenboek met onderzoeksopdrachten gebaseerd op authentieke problemen uit de psychologie</li> <li>• Boek <i>Onderzoek de Basis</i> (Van Buuren et al., 2003) met onderzoekchecklist-programma voor passende M&amp;T en data-analytische technieken</li> <li>• Boeken statistiek (Nijdam, 2003,2004)</li> <li>• just-in-time-leermaterialen die zijn toegesneden op de casussen en waarin de student op intuïtieve wijze wordt ingeleid in de desbetreffende analyse-techniek.</li> </ul>
Berekeningen	Met de hand of met gebruikmaking van rekencalculator.	Zoveel mogelijk met gebruikmaking van statistische software (SPSS versie 11).
Tentaminering	Toets met 20 onafhankelijke meerkeuzevragen (duur tentamen 3 uur) die betrekking hebben op de gehele leerstof en die vragen naar kennis en begrip; leermaterialen kunnen tijdens tentamen worden geraadpleegd. Studenten kunnen het hele jaar door wekelijks toetsen (maximum drie toetskansen).	Toets met 40 meerkeuzevragen (duur tentamen 3 uur) waarvan de ene helft van de items is gebaseerd op een casus die van tevoren door studenten met SPSS moet zijn voorbereid en de andere helft de leerstof betreft waarin wordt gevraagd naar kennis en begrip. Studenten kunnen driemaal per jaar toetsen op bepaalde momenten (maximum drie toetskansen).

### *Procesvariabelen*

#### *Attitude ten aanzien van de statistiek*

Voor de meting van de attitude ten aanzien van de statistiek is uitgegaan van de *Survey of Attitudes toward Statistics* (SATS) van Schau, Stevens, Dauphinee en Del Vecchio (1995). De SATS is een gestandaardiseerde zelfrapportageschaal bestaande uit 28 Likert-type items met antwoordcategorieën waarden die variëren van 'zeer mee oneens' tot 'zeer mee eens'. De SATS is oorspronkelijk ontworpen voor studenten die zich hebben ingeschreven voor een introductiecursus statistiek. De opbouw en samenstelling van de SATS is beschreven in tabel 5.3. De itemnummers verwijzen naar de plaats in de vragenlijst (bijlage II.2). In afwijking van de oorspronkelijke zevenpuntsindeling zijn in dit onderzoek de antwoordcategorieën naar een vijfpunts Likertschaal overgezet, waarbij met een 1 'zeer mee oneens' kan worden aangegeven en met een 5 'zeer mee eens'.

De viercomponentenstructuur van de SATS is bevestigd in een aantal onderzoeken (Dauphinee, Schau, & Stevens, 1997; Hilton, Schau & Olsen, 2004; Mills, 2004; Svah et al., 1995; Tempelaar, 2007).

TABEL 5.3 Dimensies, componenten, definities en voorbeelden van items uit de *Survey of Attitudes toward Statistics (SATS)*

<i>Dimensie</i>	<i>Compotentie</i>	<i>Definitie</i>	<i>Item voorbeeld</i>	<i>Itemnummer</i>
Affectieve dimensie	Affect	Positieve en negatieve gevoelens t.a.v. statistiek.	Ik voel me onzeker wanneer ik statistische problemen moet oplossen.	146, 147, 156, 159, 160, 166
	Waarde van statistiek	Mening over het nut, de relevantie en de waarde van statistiek in het privéleven en in een beroep.	Statistiek zou een verplicht onderdeel moeten zijn van elke academische opleiding.	150, 152, 153, 155, 157, 158, 161, 164, 170
Cognitieve dimensie	Cognitieve competentie	Mening over hun intellectuele bagage om statistiek toe te kunnen passen.	Ik kan statistiek leren.	148, 154, 165, 168, 169, 172
Gedrags- dimensie (conatie)	Moeilijkheid	Mening over de moeilijkheid van het vak (de cursus) statistiek.	Statistische formules zijn gemakkelijk te begrijpen.	149, 151, 162, 163, 167, 171, 173

#### *Motivatatie*

Motivatatie is gemeten met items uit het motivatiedeel van de *Motivated Strategies for Learning Questionnaire* (MSLQ, Pintrich, Smith, Garcia, & McKeachie, 1991). De volledige MSLQ is een gestandaardiseerde zelfrapportageschaal bestaande uit 81 zevenpunts Likertschaalitems. Het motivatiedeel van de MSLQ beslaat 31 items die de vier componenten Intrinsieke waarde, Taakwaardering, Zelfeffectiviteit en Testangst vormen. De opbouw en samenstelling van de schalen en items is beschreven in tabel 5.4. Voor dit onderzoek zijn de antwoordcategorieën naar een vijfpunts Likertschaal overgezet, waarbij met een 1 'zeer mee oneens' kan worden aangegeven en met een 5 'zeer mee eens'.

De MSLQ is gevalideerd in een groot aantal onderzoeken, variërend van leerlingen op basisscholen tot studenten op universiteiten (Midgley, Kaplan, Middleton, & Maehr 1998; Pintrich, Smith, Garcia, & McKeachie, 1993). Ook is deze schaal in veel landen en in verschillende disciplines gebruikt (Chen, 2002).

TABEL 5.4 Dimensies, componenten, definities en voorbeelden van items uit het motivatiegedeelte van de *Motivated Strategies for Learning Questionnaire* (MSLQ)

Dimensie	Component	Definitie	Item voorbeeld	Itemnummer
Waarde	Intrinsieke waarde (Waarom doe ik dit?)	Reden om een taak aan te pakken, daarbij betrokken te zijn en een prestatie te leveren als een doel op zich.	In deze modules put ik de meeste voldoening uit het begrijpen van methoden en statistiek voor het toepassen van onderzoek.	4, 5, 6, 7, 8, 76, 77, 78, 79
	Taakwaardering (Wat vind ik van deze taak?)	Evaluaties over het belang, het nut en de relevantie van de taak.	Ik ben zeer geïnteresseerd in het inhoudelijke aspect van deze cursus(sen).	9, 10, 11, 12, 13, 14
Verwachting	Zelfeffectiviteit (Kan ik deze taak uitvoeren?)	Oordeel over de eigen capaciteiten om een taak uit te voeren.	Ik ben er zeker van dat ik zelfs de moeilijkste onderwerpen binnen de leerstof van deze cursus(sen) kan begrijpen.	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26
Testangst	Testangst (Als ik maar niet zak?)	Zorgen en een negatief gevoel hebben over de eigen prestaties	Ik heb een ongemakkelijk en ellendig gevoel als ik tentamen doe.	27, 28, 29, 30

### Leerstrategieën

Leerstrategieën zijn gemeten met het 'strategies for learning'-gedeelte van de *Motivated Strategies for Learning Questionnaire* (MSLQ) (Pintrich et al., 1991). Het gedeelte van deze schaal over leerstrategieën bestaat uit 31 items die betrekking hebben op de inzet van verschillende cognitieve strategieën (Herhalen, Elaboreren, Organiseren, Kritisch denken) en metacognitieve strategieën (Zelfregulatie) bij het leren. De opbouw en samenstelling van de schalen en items is beschreven in tabel 5.5. De oorspronkelijke zevenpunts Likertschaalitems zijn omgezet naar een vijfpunts schaal met antwoordcategorieën variërend van 1 ('zeer mee oneens') tot 5 ('zeer mee eens').

TABEL 5.5 Dimensies, componenten, definities en voorbeelden van gebruikte items uit het gedeelte betreffende het leren van de *Motivated Strategies for Learning Questionnaire* (MSLQ)

<i>Dimensie</i>	<i>Component</i>	<i>Definitie</i>	<i>Item voorbeeld</i>	<i>Itemnummer</i>
Cognitieve strategie	Herhalen	Het opzeggen of noemen van onderwerpen (items) uit een lijst die van buiten moet worden geleerd (om de aandachts- en inslijpingsprocessen (ecoding) te beïnvloeden zonder de kenniselementen te verbinden of de informatie met aanwezige voorkennis te integreren).	Als ik me voorbereid op een tentamen, dan zeg ik de feiten uit de theorie voor mezelf steeds hardop.	31, 32, 33, 75
	Elaboreren	Het opslaan van informatie in het langetermijngeheugen door interne verbindingen te maken tussen onderwerpen die moeten worden geleerd en deze te integreren en verbinden met de reeds aanwezige voorkennis (door parafraseren, opsommen en maken van analogieën die reeds geleerd zijn).	Als ik een onderwerp bestudeer, probeer ik dat met informatie uit verschillende bronnen te verbinden.	34, 35, 36, 37, 38, 39, 40, 41
	Organiseren	Het selecteren van geschikte informatie en het leggen van verbindingen tussen de verschillende informatie-elementen die moeten worden gekend (om zeer betrokken te raken bij de taak en zo beter te kunnen presteren (door samen te voegen ('clustering'), maken van samenvattingen en selecteren van de voornaamste ideeën in leesteksten).	Als hulp bij het studeren maak ik een samenvatting van de kernpunten uit het materiaal.	43, 44, 45, 46, 47, 48, 49
	Kritisch denken	Het maken van kritische beoordelingen door eerder bestudeerde kennis in nieuwe situaties toe te passen (om problemen op te lossen, beslissingen te nemen of om kritische beoordelingen te maken m.b.t. hoge standaarden)	Bij het lezen van onderzoeksverslagen kijk ik altijd meteen naar de grootte van de onderzoeksgroep om te zien of de conclusies daarvoor wel voldoende ondersteund worden.	50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61



<i>Dimensie</i>	<i>Component</i>	<i>Definitie</i>	<i>Item voorbeeld</i>	<i>Item nr</i>
Meta-cognitieve strategie	Zelfregulatie (verwijst naar de gewaarwording van, de kennis van en de controle over cognities)	Plannen: het stellen van doelen en het analyseren van de taak (om voorkennis te activeren, waardoor het gemakkelijker wordt om de leerstof te organiseren en te begrijpen).	Voordat ik met studeren begin, denk ik na over de dingen die ik wil gaan doornemen.	63, 64, 65, 66, 67, 68, 69, 70, 71, 72 73, 74
		Monitoren: het sturen en richten van de aandacht en het zichzelf testen door het stellen van vragen. (om het begrijpen en het doorgronden van de leerstof en de integratie met voorkennis te ondersteunen).	Ik stel mezelf vragen om er zeker van te zijn dat ik het materiaal dat ik heb bestudeerd, beheers.	
		Regulatie: het afstemmen en continu evalueren van de eigen cognitieve activiteiten. (Regulerende activiteiten verbeteren het leerproces omdat ze de student ondersteunen bij het checken en het corrigeren van het leergedrag bij het uitvoeren van de leertaak.)	Als het studiemateriaal moeilijk is, geef ik het óf op óf ik bestudeer alleen de eenvoudige delen.	

*Productvariabelen: leeruitkomsten*

## Functionele kennis van de statistiek

Voor het aantonen van de beheersing van functionele kennis van de statistiek is een statistiekkennistoets ontwikkeld bestaande uit negen vraagstukken (zie tabel 5.6; zie tevens de vragenlijst in bijlage II.2, items 137-145). De vragen hebben betrekking op een aantal min of meer 'alledaagse' voorbeelden die te maken hebben met het doen van onderzoek en de statistische verwerking van gegevens. Studenten is gevraagd hun antwoorden in eigen woorden zo uitgebreid mogelijk toe te lichten. Wellicht ten overvloede is de studenten gemeld dat ze niet op hun antwoorden beoordeeld zouden worden. Op deze manier is geprobeerd te bewerkstelligen dat studenten op een testangstvrije manier zo veel mogelijk van hun kennen en kunnen op het terrein van de statistiek blij kunnen geven.

De antwoorden op de open vragen zijn door twee onafhankelijke beoordelaars volgens de SOLO-taxonomie (*Structure of the Observed Learning Outcomes*, Biggs & Collis, 1982) naar mate van functionele kennis van de statistiek geclassificeerd. Cohen's Kappa van de intercodeursbetrouwbaarheid bedroeg voor de negen items tussen de .90 en 1.0. De SOLO-taxonomie onderscheidt vijf niveaus van begrip, van 'niet weten' (pre-structureel, niveau E) tot 'expert'-niveau (extended abstract, niveau A) (zie tabel 5.7). SOLO hanteert als niveaus de letters A t/m E. In dit onderzoek zijn deze aanduidingen

vertaald in  $A = 4$ ,  $B = 3$ ,  $C = 2$ ,  $D = 1$  en  $E = 0$ . Het gemiddelde van de toegekende scores over de items geeft het niveau van de statistiekkennis weer.

TABEL 5.6 Opgaven statistiekkennis

Type vraag	Omschrijving	Itemnummer
Invulvraag	De vraag is gebaseerd op Schau en Mattern's (1997) visueel-ruimtelijk netwerk (ook wel met 'concept map' of 'mind map' aangeduid) van met elkaar verbonden statistische begrippen. Dit netwerk van begrippen representeert de cognitieve structuur die 'experts' onderscheidt van 'novices' (leken). Het onderscheid daartussen speelt vooral in de mate van complexiteit waarin zo'n schema of een verzameling van onderling verbonden schema's is uitgewerkt, alsmede de nauwkeurigheid en de relevantie van de opgenomen begrippen (Marshall, 1995). De kaart bevat begrippen (die als knooppunt in een kaart fungeren en worden gerepresenteerd door ovalen) en de verbindingen daartussen (gerepresenteerd met pijlen die naar de te verbinden knooppunten (begrippen) wijzen). In de kaart stellen de knooppunten en de pijlen het proces voor van het toetsen van hypothesen over relaties tussen variabelen en over verschillen tussen groepen. Zeven ovalen zijn leeg. Studenten dienen de conceptmap te completeren door in de lege ovalen begrippen te plaatsen die ze kunnen kiezen uit een toegevoegd lijstje met begrippen.	137
Meerkeuze vraag	Met deze vraag (eveneens afkomstig van Schau en Mattern, 1997) wordt kennis en begrip van onderlinge verbanden getoetst, in dit geval met de standaardafwijking als uitgangspunt.	138
Open	Studenten staan voor de taak de uitkomsten van een onderzoek in eigen woorden 'terug te vertalen' naar drie mogelijk ten grondslag liggende onderzoeksvragen en hypothesen.	139
Open	Studenten wordt gevraagd om een uitgebreide invulling te geven van de vuistregel voor de standaardafwijking.	140
Open	Studenten wordt gevraagd een misconceptie op te helderen. De misconceptie betreft een uitleg van een doctoraalstudent psychologie aan de begeleider hoe de resultaten van een onderzoek moeten worden geïnterpreteerd. Uit de uiteenzetting blijkt dat de kandidaat geen idee heeft van de betekenis van correlatie, statistische significantie, proportie verklaarde variantie en overschrijdingskansen ( $p$ -waarde).	141
Open	In deze vraag wordt een pijlendiagram gepresenteerd van een relatie tussen twee variabelen die wordt gemodereerd door een derde (dichotome) variabele. Voor elk van de twee categorieën van de derde variabelen wordt een puntendiagram tussen de twee variabelen getoond. Studenten is gevraagd argumenten en interpretaties te geven om de betekenis en de relatie van het pijlendiagram met de twee puntendiagrammen te duiden.	142
Meerkeuze vraag	Gegeven enkele statistische uitkomsten betreffende een beleidsprobleem wordt studenten gevraagd aan te geven welke van de twee voorgestelde beleidsmaatregelen in overeenstemming is met de onderzoeksuitkomsten.	143
Open	Overgenomen van Reading (1996, 1999). Dit item betreft een grafiek van een frequentieverdeling (polygoon). Studenten wordt gevraagd deze grafiek in termen van univariate statistische maten te beschrijven.	144
Combinatie van meerkeuze en open	Deze opgave betreft een taak met open en gesloten vragen. Deze hebben betrekking op (1) aspecten van het onderzoeksontwerp, (2) theoretische aspecten, (3) aspecten rondom onderzoeksprocedures en (4) niet-technische aspecten van onderzoek. Dit item is gebaseerd op studies naar effecten van het gebruik van de onderzoeksscenario's waarin kennisstructuren betreffende de inferentiële statistiek van experts en novices worden vergeleken (Alacaci, 2004).	145

TABEL 5.7 *Structure of Observed Learning Outcomes (SOLO) Taxonomy (Biggs & Collis, 1982, 1989)*

<i>SOLO-niveau</i>	<i>SOLO-indicator</i>	<i>Beschrijving*</i>
E Prestructureel	mist het punt	'An outcome reflecting student's engagement (Biggs & Collis, 1989), but irrelevant to the topic in question (Trigwell, 1999), which shows that nothing at all is known about the area under consideration (Boulton-Lewis, 1994) or which elicits no meaningful response (Ramsden, 1992). An idiosyncratic response (Watson et al., 2003).' ( <i>In het geheel geen kennis en dus ook geen begrip</i> )
D Unistruktuur	noemen van een relevant aspect	'Outcomes are relevant (Tang, 1998), but contain oversimplified generalizations of only one aspect of the topic (Biggs, 1979; Dahlgren, 1997).' ( <i>Begrip is nominaal</i> )
C Multistruktuur	opsommen van relevante aspecten	'Two or more relevant items are included, but are presented independently (Prosser & Trigwell, 1999) and not linked or integrated with each other (Boulton-Lewis & Dart, 1994). Outcomes at this level involve selective and premature closure (Biggs, 1979).' ( <i>Begrip als er iets van weten</i> )
B Relatieel	verbinden van items in een lijst	'The essential difference between this level and the previous one is that the learned information is integrated (Biggs & Collis, 1989), leading to presentation of a meaningful, coherent and whole structure (Hattie & Purdie, 1998; Tang, 1998). Biggs (1979) states that, for inclusion at this level, in addition to the emphasis on meaningful integration, most or all of the relevant knowledge elements should be included. However, Prosser & Trigwell (1999) state that outcomes at this level do not necessarily include a greater amount of content than in the previous level and focus more on the issues of integration and coherence.' ( <i>Begrip van de onderliggende relaties</i> )
A Uitgebreid abstract	integreren van ideeën in begrippen in een discipline	'Trigwell (1999) describes this as the most "complete" level. Outcomes demonstrate that understanding of the integrated knowledge can be generalized or transferred to new situations or experiences (Dahlgren, 1997; Boulton-Lewis, 1998) and higher levels of abstraction (Hattie & Purdie, 1998). Biggs (1997) also states that this level is characterized by questioning of basic arguments and existing knowledge, meaning that closure is regarded as inappropriate.' ( <i>Begrip in de vorm van verre transfer met inbegrip van metacognitieve kennis</i> )

\* Beschrijvingen zijn afkomstig van Morris (2001, p. 504).

Volgens Veldhuis-Diermanse (2002) is de SOLO-taxonometrie een krachtig instrument om de kennisstructuren en het gevormde begrip van de leerstof bij studenten in het hoger onderwijs te analyseren. Empirisch onderzoek in een aantal disciplines heeft aangetoond dat de taxonomie contextonafhankelijk is (Callingham & Watson, 2005; Campbell, Watson & Collis, 1992; Chan, Tsui & Chan, 2002; Holmes, 2005; Pfannkuch, 1999, 2005; Reading, 1996, 2004; Scholten, Keeses & Lawson, 2002; Watson & Callingham, 2003; Watson & Moritz, 2000, 2001). De taxonomie is onder andere gebruikt bij de interpretatie van de cognitieve ontwikkeling in een verscheidenheid

aan leergebieden zoals kritische gecijferdheid (Lake, 1999), het conceptualiseren en construeren van essays (Campbell et al., 1998), advisering (Burnett, 1999), breuken, (Watson, Campbell & Collis, 1996), meetkunde (Pegg, 1997), wetenschappelijke begrippen (Panizzon & Pegg, 1997), 'statistical literacy' (Watson & Moritz, 2000), statistisch redeneren (Reading, 1996, 1998, 2000), statistisch denken (Jones, Langrall, Thornton, & Mogill 1997), sliks patronen (Scholten et al., 2002) en het construeren van kennis in asynchrone CSCL-omgevingen (Veldhuis-Diermanse, 2002).

#### *Autonomie*

Autonomie is in dit onderzoek gedefinieerd als de mate waarin iemand zelfstandig een onderzoek uit kan voeren. Autonomie wordt gemeten met negen items die zijn gebaseerd op de 'intellect/autonomie-factor' van de Five-Factor Personality Inventory ('Big Five', Goldberg, 1990). De items zijn aangepast in de context van het zelfstandig uitvoeren van onderzoek. Een voorbeeld van een autonomie-item luidt: 'Ook al loop ik tegen onvoorziene problemen op in mijn onderzoek, dan bedenk ik daar toch een goede oplossing voor.' Studenten geven op een vijf puntsschaal met antwoordcategorieën variërend van (1) 'helemaal niet van toepassing' tot (5) 'helemaal van toepassing' aan in hoeverre ze het eens zijn met de stelling. De autonomieschaal in de vragenlijst betreft de volgende items: 113, 114, 116, 117, 119, 120, 124, 127, 131.

#### *Afhankelijkheid*

Afhankelijkheid is in dit onderzoek gedefinieerd als de behoefte aan van boven opgelegde structuur en sturing, gebrek aan voldoende zelfeffectiviteit en behoefte aan sociale ondersteuning. Afhankelijkheid wordt gemeten met tien items die eveneens zijn gebaseerd op de Big Five (Goldberg, 1990). Een voorbeeld van een item is: 'Bij onderzoek doen houd ik me altijd letterlijk aan de opdracht zoals die in het werkboek is omschreven, dan ben ik er zeker van dat ik het doe zoals het hoort.' Studenten geven op een vijf puntsschaal met antwoordcategorieën variërend van (1) 'helemaal niet van toepassing' tot (5) 'helemaal van toepassing' aan in hoeverre ze het eens zijn met de stelling. De afhankelijkheidsschaal in de vragenlijst bestaat uit de volgende items: 115, 118, 121, 122, 123, 125, 126, 128, 129, 130.

### 5.2.3 *Procedure*

Voorafgaand aan de verzending van de vragenlijst is studenten een brief gestuurd waarin het belang van het onderzoek voor zowel studenten als voor de opleiding Psychologie is uiteengezet en waarin om hun medewerking is verzocht. Na twee weken kregen de studenten de vragenlijst toegestuurd. Studenten waarvan een e-mailadres bekend was, kregen een verwijzing naar een URL op het internet. Via dit adres kon de vragenlijst online worden ingevuld en worden geretourneerd. Studenten waarvan geen e-mailadres bekend was, kregen de vragenlijst per post toegestuurd (met een instructie voor het invullen van de vragenlijst, en een portvrije antwoortenveloppe). Twee weken nadat de vragenlijsten waren verzonden, kregen studenten een eerste herinnering toegezonden. In deze herinnering werd nogmaals gewezen op het belang

van het onderzoek. Na een maand werden studenten die nog niet hadden gereageerd, telefonisch benaderd. Na twee maanden is een aanvullend onderzoek gestart naar de nonrespons. Voor een verslag daarvan wordt verwezen naar bijlage III.1.

#### 5.2.4 *Analyseschema*

Voorafgaand aan de toetsing van de hypothesen zijn de gebruikte schalen en de test voor de statistiekennis geanalyseerd volgens de procedures van het Raschmodel (Bond & Fox, 2001; Linacre, 2005). Voor deze Raschmodeloefening zijn twee redenen te noemen. De eerste is om schalen te verkrijgen van intervalmeetniveau. De tweede is om strenger dan met de interne consistentieprocedure met confirmatorische factoranalyse (CFA) mogelijk is, te toetsen op eendimensionaliteit. Beide aanpassingen – meting op intervalniveau en eendimensionaliteit – zijn nodig om aan de voorwaarden voor analyses met MANOVA (alleen meetniveau) en Structural Equation Modeling (SEM) te voldoen. Daarnaast biedt Raschanalyse meer waarborgen voor betrouwbaarheid en validiteit i.c. constructvaliditeit van de meetinstrumenten (Bond & Fox, 2001; Curda, 1997; Ter Laak & De Goede, 2003). Het opnemen van twee belangrijke doch deels verwante psychologische constructen in het determinantenmodel, te weten: attitude en motivatie, zou een bron van ruis in de analyses kunnen zijn (cross-loadings).

Voor een beschrijving van de voorstudie met de Raschmodelanalyses, de criteria en een overzicht van de meetinstrumenten wordt verwezen naar bijlage III.2. Voor de Raschanalyses is gebruik gemaakt van het programma Winsteps versie 3.6 (Linacre, 2005). Op één meetinstrument na, te weten: Organiseren, hebben alle gebruikte Likert-schalen de Raschanalyses ‘overleefd’, hoewel soms ingrijpende aanpassingen hebben moeten plaatsvinden. Van de 107 items zijn er 13 items uit de diverse meetinstrumenten verwijderd en zijn er 23 items in een andere schaal terechtgekomen.

In de resultatenparagraaf zullen eerst de effecten van de twee onderwijsbenaderingen (hypothesen 1 t/m 5) door middel van vergelijking van gemiddelden en effectmaten via MANOVA en afzonderlijke ANOVA's worden beschreven (zie paragraaf 5.3.1). Vervolgens zal in paragraaf 5.3.2 met SEM-meer-groepenanalyse (Byrne, 2001) worden nagegaan in hoeverre de procesvariabelen de relatie tussen de leeromgeving en de leeruitkomsten mediëren (hypothese 6). Volgens Cole en Maxwell (2003) volstaat het gebruik van crossectionele data niet om mediatie aan te tonen. Voor het aantonen van mediërende processen zijn minstens ‘three waves’ aan data nodig zodat relaties vastgelegd kunnen worden tussen de onafhankelijke variabele(n) op tijdstip 1, de mediërende variabele(n) op tijdstip 2 en de uitkomstvariabele(n) op tijdstip 3. Helaas biedt het huidige onderzoek slechts crossectionele data. Dit betekent dat alleen een tentatieve verkenning mogelijk is naar aanwijzingen voor mediatie. De kruisvalidatie van het model met SEM is beschreven in bijlage III.3. Voor de mediatieanalyses met SEM is gebruik gemaakt van het pakket AMOS versie 5 (Arbuckle, 2003).

### 5.3 Resultaten

#### 5.3.1 *Vergelijking van de twee leeromgevingen*

In de hypothesen 1 t/m 5 is gepostuleerd dat het competentiegericht statistiekonderwijs ten opzichte van het vakgerichte statistiekonderwijs *positieve* effecten heeft. Verwacht wordt dat degenen die competentiegericht onderwijs volgen een positievere attitude, een sterkere motivatie, meer diepgaande leersstrategieën meer functionele kennis van de statistiek, meer autonomie en minder afhankelijkheid hebben dan degenen die het vakgerichte onderwijs volgen. Deze hypothesen zijn nog algemeen geformuleerd. Gespecificeerd in deelhypothesen die voor elke component de richting van het effect benoemt, betekent dit dat er in COMP hogere gemiddelde scores verwacht worden op Affect, Waarde van statistiek, Cognitieve competentie, Intrinsieke waarde, Taakwaardering, Zelfeffectiviteit, Elaboreren, Kritisch denken, Zelfregulatie, Autonomie en Statistiekennis en lagere gemiddelde scores op Moeilijkheid, Testangst, Herhalen en Afhankelijkheid dan in VAK. Om hierover uitsluitsel te verkrijgen is een MANOVA uitgevoerd gevolgd met afzonderlijke ANOVA's per afhankelijke variabele. Aan de assumpties om deze variantie-analytische technieken te mogen toepassen, is bij alle variabelen voldaan (normaliteit, lineariteit, uitschieters, homogeïteit van variantie- en covariantiematrices en multicollineariteit). Tabel 5.8 geeft de verschillen tussen de gemiddelden op de schalen in COMP en VAK weer. De gemiddelden en standaardafwijkingen zijn uitgedrukt in logits volgens het Raschmodel. Een logit is een meeteenheid die resulteert na transformatie van ruwe scores naar log odds ratios. In de praktijk variëren de logits van - 4 ('gemakkelijk') tot + 4 ('moeilijk'). De waarde 0.0 logits wordt toegekend aan het gemiddelde van de moeilijkheidswaarden van de items.

MANOVA laat een significant verschil zien tussen de twee leeromgevingen op de combinatie van afhankelijke variabelen ( $F(15, 452) = 4.616, p = .000$ ); Wilks' Lambda is .867 en de partiële  $\eta^2$  is .13). De effectgrootte is volgens Cohen (1988) te karakteriseren als een middelmatig effect.

Worden vervolgens als exploratieve follow-up met ANOVA's de verschillen per variabele bekeken, dan blijkt met uitzondering van Herhalen de gemiddelden volgens verwachting te verschillen. Echter, de verschillen op Waarde van statistiek, Elaboreren en Afhankelijkheid zijn niet statistisch significant. Wat betekent dit voor de hypothesen 1 t/m 6?

Ten aanzien van hypothese 1 kan gesteld worden dat de deelhypothesen die betrekking hebben op de afzonderlijke attitudecomponenten in COMP volgens verwachting positiever zijn dan in VAK. Op de deelhypothese betreffende Waarde van statistiek na, zijn deze bevestigd. Vanwege het niet significant zijn van het verschil bij Waarde van statistiek moet deze deelhypothese worden verworpen.

TABEL 5.8 ANOVA's gevolgd op MANOVA voor verschillen in gemiddelden (in logits) tussen COMP ( $n = 128$ ) en VAK ( $n = 340$ )

	<i>M</i> COMP	<i>SD</i> COMP	<i>M</i> VAK	<i>SD</i> VAK	<i>F</i>	<i>df</i>	<i>Sig</i>	<i>Part</i> <i>eta</i> <sup>2</sup>
<i>Attitude</i>								
Affect	0.67	2.03	-0.13	2.12	13.264	1, 467	.000	.028
CognCompetentie	1.47	1.57	0.74	2.08	13.121	1, 467	.000	.027
WaardeStatistiek	0.65	1.48	0.42	1.76	1.687	1, 467	.195	.004
Moeilijkheid	0.32	1.19	0.77	1.52	9.048	1, 467	.003	.019
<i>Motivatie</i>								
IntrinsiekeWaarde	2.57	3.22	1.21	3.27	16.200	1, 467	.000	.034
Taakwaardering	1.96	1.76	1.29	1.70	14.037	1, 467	.000	.029
Zelfeffectiviteit	1.19	1.93	0.35	2.78	9.837	1, 467	.002	.021
Testangst	-2.82	2.99	-2.06	3.37	5.090	1, 467	.026	.011
<i>Leerstrategie</i>								
Herhalen	-0.26	2.09	-0.72	1.96	4.988	1, 467	.026	.011
Elaboreren	1.85	1.87	1.69	1.76	0.712	1, 467	.399	.002
KritischDenken	1.20	1.57	0.73	1.40	9.889	1, 467	.002	.021
Zelfregulatie	1.10	0.90	0.74	0.88	15.311	1, 467	.000	.032
<i>Leeruitkomsten</i>								
Autonomie	1.41	1.40	0.73	1.49	19.948	1, 467	.000	.041
Afhankelijkheid	-0.71	1.79	-0.71	1.66	0.000	1, 467	.987	.000
Statistiekkennis	-2.21	2.09	-2.98	1.95	13.931	1, 467	.000	.029

Hypothese 2 die uiteengelegd kan worden in deelhypothesen betreffende Intrinsieke waarde, Taakwaardering, Zelfeffectiviteit en Testangst wordt voor alle motivatiecomponenten bevestigd. De gemiddelden voor Intrinsieke waarde, Taakwaardering en Zelfeffectiviteit zijn alle volgens verwachting hoger en voor Testangst lager in COMP dan in VAK.

Hypothese 3 bevat de deelhypothesen omtrent de leerstrategieën Herhalen, Elaboreren, Kritisch denken en Zelfregulatie. In tegenstelling tot de verwachting bij Herhalen blijkt de gemiddelde score in COMP *hoger* te zijn dan in VAK. De deelhypothese voor Herhalen wordt hiermee verworpen. Ook de deelhypothese voor Elaboreren wordt verworpen vanwege het niet-significante verschil tussen beide groepen.

De drie hypothesen over de leeruitkomsten blijken alle drie volgens verwachting uit te vallen. Echter, de hypothese betreffende Afhankelijkheid moet worden verworpen aangezien het verschil in beide groepen niet statistisch significant blijkt.

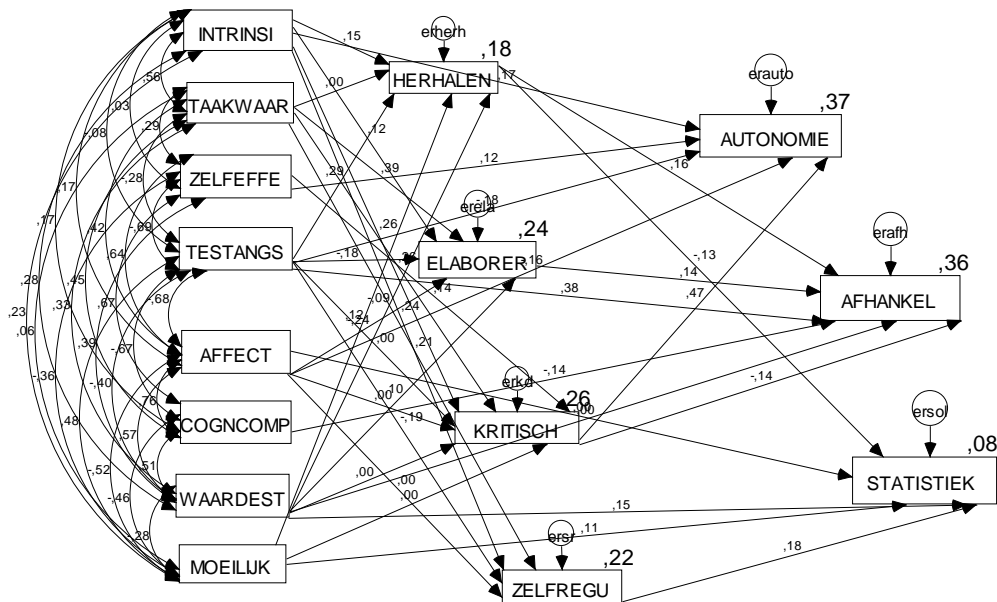
*Samengevat:* Op de deelhypothese voor Herhalen na zijn de verschillen in gemiddelde scores tussen COMP en VAK alle volgens verwachting. Wel moeten de deelhypothesen over Waarde van statistiek, Elaboreren en Afhankelijkheid vanwege niet significante verschillen verworpen worden.



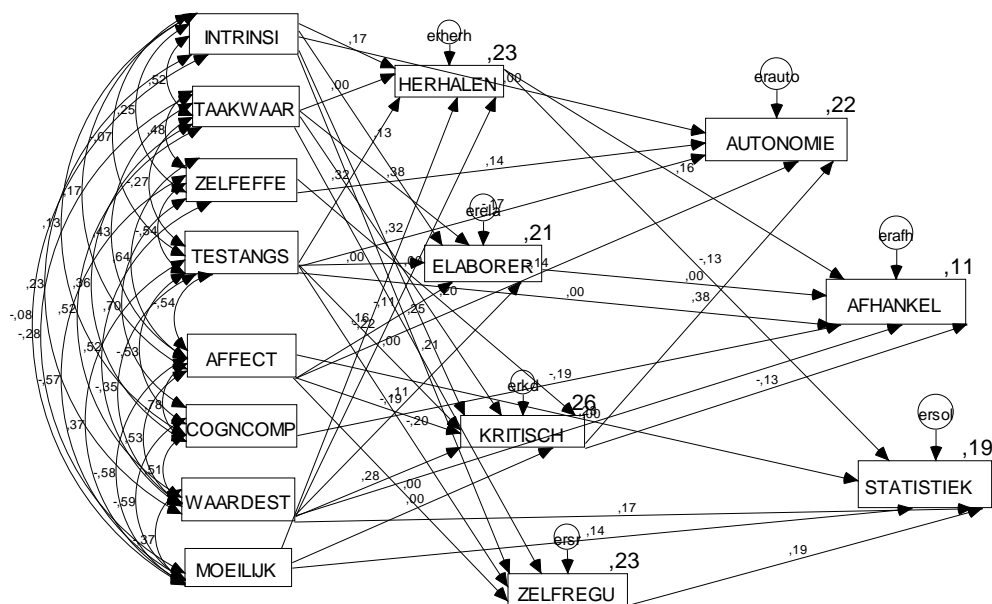
## 5.3.2 Toetsing van de mediatie in het model met SEM

Volgens het determinantenmodel mediëren attitude, motivatie en leerstrategieën als procesvariabelen de relatie tussen leeromgeving en leeruitkomsten. Deze procesvariabelen worden beschouwd als causale mechanismen waardoor studenten tot bepaalde leeruitkomsten komen. Hoewel zoals eerder gesteld de toetsing een tentatief karakter zal hebben, wordt eerst de validiteit van de structuur van het model onderzocht. De kruisvalidatie ('cross validation', Byrne, 2001) is uitgevoerd op het gecombineerde databestand van de twee groepen VAK en COMP. Na samengevoeging is het bestand volgens toeval weer opgedeeld in twee even grote steekproeven ( $n = 234$  respondenten). In de calibratiesteekproef zijn de variabelen en de relaties daartussen op invariantie van de causale structuur getoetst en vervolgens in de validatiesteekproef gerepliceerd. Voor de kruisvalidatie met SEM is gebruik gemaakt van het pakket AMOS versie 5 (Arbuckle, 2003). Voor een beschrijving van de kruisvalidatie wordt verwezen naar bijlage III.3. Figuur 5.2 toont het structurele model (padmodel) dat als resultaat van de kruisvalidatietoets tevoorschijn is gekomen. Voor de validatiesteekproef bedragen de fitmaten van het uiteindelijke model  $\chi^2 = 87.11$  met 55 vrijheidsgraden ( $p = .00$ );  $GFI = .95$ ;  $CFI = .97$ ;  $RMSEA = .06$ .

De resultaten die met meer-groepenanalyse binnen SEM (Multi-group Analysis, Byrne, 2001) zijn verkregen, zijn weergegeven in de figuren 5.2A en B en in de tabellen 5.9A en B. Deze tabellen geven de significante, gestandaardiseerde padcoëfficiënten ( $\beta$ ) en proporties verklaarde varianties ( $R^2$ ) weer voor de modellen in COMP en VAK. Op de verschillende causale paden in het model wordt nu ingegaan.



FIGUUR 5.2A Geoperationaliseerd determinantenmodel COMP



FIGUUR 5.2B Geoperationaliseerd determinantenmodel VAK

De goodness of-fit van het constrained gecombineerde model is als volgt:  $\chi^2 = 234.64$ ,  $df = 135$ ,  $p = .00$ ;  $GFI = .94$ ;  $CFI = .96$ ;  $TLI = .94$ ;  $RMSEA = .04$  [.03 - .05].

### Leerstrategieën

Zowel in COMP als in VAK domineren de motivatiecomponenten de significante regressiepaden naar de leerstrategieën. In COMP zijn er elf significante paden vanuit motivatiecomponenten naar leerstrategieën, met slechts vier vanuit de attitudecomponenten. In VAK manifesteren zich negen significante regressiepaden vanuit motivatiecomponenten naar de leerstrategieën en zes vanuit de attitudecomponenten. In beide groepen kan het overgrote deel van de proportie verklaarde variantie in de leerstrategieën worden toegeschreven aan de motivatiecomponenten.

Een opvallend verschil tussen het regressiepadpatroon binnen COMP en VAK is dat in COMP de proportie verklaarde variantie in Kritisch denken (26%) geheel voor rekening komt van de motivatiecomponenten, terwijl in VAK de attitudecomponenten Affect en Waarde van statistiek samen een deel daarvan (10% van de 26%) voor hun rekening nemen.

In beide groepen zijn alleen de motivatiecomponenten verantwoordelijk voor de proportie verklaarde variantie in Zelfregulatie.

TABEL 5.9A Gestandaardiseerde padcoëfficiënten ( $\beta$ ), overschrijdingskansen  $p$  en  $R^2$  waarden van het determinantenmodel binnen COMP

Naar →	Herhalen			Elaboreren			Kritisch denken			Zelfregulatie			Autonomie			Afhankelijkheid			Statistiek-kennis			
Van ↑	β	p	R²	β	p	R²	β	p	R²	β	p	R²	β	p	R²	β	p	R²	β	p	R²	
Intrinsieke Waarde	.15	.00	.02	.12	.01	.02	.26	.00	.10	.24	.00	.08	.17	.02	.04							
Taakwaardering				.39	.00	.19	.26	.00	.16	.21	.00	.09										
Zelfeffectiviteit							.14	.00	-				.12	.02	.02							
Testangst	.29	.00	.09	-.18	.04	.02				-.19	.00	.05	-.18	.00	.04	.38	.00	.29			-	
Motivatie				.11		.23			.26			.22			.10			.29				
Affect				-.24	.00	-							-.16	.01	-							
Cognitieve Competentie Waarde																-.15	.00	.01		.15	.00	.03
Statistiek	-.09	.02	.02	.10	.03	.01													.11	.01		
Moeilijkheid	.12	.00	.05																			
Attitude				.07		.01			-			-			.00			.01			.03	
Herhalen																.16	.00	.03	-.14	.00	.02	
Elaboreren																.15	.04	.01				
Kritisch denken													.47	.00	.27	-.14	.00	.02				
Zelfregulatie																			.18	.00	.03	
Leerstrategieën															.27			.06			.05	
Totaal			.18			.24			.26			.22			.37			.36			.08	

*Leeruitkomsten*

Eerst wordt aandacht besteed aan de bevindingen over de statistiekkennis. Daarna wordt ingegaan op de resultaten over autonomie en afhankelijkheid.

Het meest opvallende verschil in het regressiepadpatroon richting de statistiekkennis tussen COMP en VAK is dat in COMP een significant regressiepad vanaf Affect ontbreekt. Het pad van de attitudecomponent Affect naar statistiekkennis blijkt uniek voor VAK. In COMP wordt de meeste variantie in statistiekkennis verklaard door de leerstrategieën Zelfregulatie en Herhalen (5% van de 8%) en de overige 3% door de attitudecomponent Waarde van statistiek. Moeilijkheid heeft in COMP weliswaar eensignificante bèta maar verklaart geen unieke variantie in statistiekkennis. In tegenstelling met COMP wordt in VAK de meeste variantie in statistiekkennis verklaard door Affect (11% van 19%).

TABEL 5.9B Gestandaardiseerde padcoëfficiënten ( $\beta$ ), overschrijdingskansen  $p$  en  $R^2$  waarden van het theoretisch model binnen VAK

Naar →	Herhalen			Elaboreren			Kritisch denken			Zelfregulatie			Autonomie			Afhankelijkheid			Statistiekennis		
	$\beta$	$p$	$R^2$	$\beta$	$p$	$R^2$	$\beta$	$p$	$R^2$	$\beta$	$p$	$R^2$	$\beta$	$p$	$R^2$	$\beta$	$p$	$R^2$	$\beta$	$p$	$R^2$
Van ↑																					
Intrinsieke Waarde	.17	.00	.01	.13	.01	.03	.32	.00	.15	.25	.00	.12									
Taakwaardering				.38	.00	.15				.21	.00	.06									
Zelfeffectiviteit							.20	.00	.01				.14	.02	.01						
Testangst	.32	.00	.17							-.20	.00	.05	-.17	.00	.03						
Motivatie				.18			.18			.16			.23			.04					
Affect				-.22	.00	.01	-.19	.00	.02				-.14	.01	-				.23	.00	.11
Cognitieve Competentie Waarde																-.19	.00	.07			
Statistiek	-.11	.02	.01	.11	.03	.02	.28	.00	.08										.17	.00	.02
Moeilijkheid	.16	.00	.04																.14	.01	.01
Attitude				.05			.03		.10				-			.07					.14
Herhalen																.16	.00	.02	-.13	.00	.02
Elaboreren																					
Kritisch denken													.38	.00	.18	-.13	.00	.02			
Zelfregulatie																			.19	.00	.03
Leerstrategieën																.18		.04			.05
Totaal			.23			.21			.26			.23			.22			.11			.19

Drie attitudecomponenten (Affect, Waarde van statistiek en Moeilijkheid) verklaren in VAK 14% van de 19% aan verklaarde variantie in statistiekkennis. Net als in COMP nemen de leerstrategieën Zelfregulatie en Herhalen samen 5% verklaarde variantie voor hun rekening. Echter, in VAK domineert het pad van Affect naar statistiekkennis het regressiepadpatroon. Moeilijkheid blijkt in beide groepen een positief regressiepad te vormen. Dit komt overeen met onderzoeksbevindingen dat taken een uitdaging moeten zijn. De statistiekkennis lijkt in VAK te staan of vallen met de attitude ten aanzien van statistiek. In COMP hebben de leerstrategieën meer impact op de statistiekkennis. Zelfregulatie, dat vooral wordt bepaald door motivatie, levert daaraan de meest positieve bijdrage.

De leerstrategie Kritisch denken blijkt in COMP en VAK de krachtigste predictor van Autonomie. Echter, in COMP is er naast een significant negatief pad vanaf Testangst naar Autonomie tevens een vrijwel even krachtig significant positief pad richting Autono-

mie vanuit Intrinsieke waarde. Dit pad is uniek voor COMP. Daarnaast vertoont COMP twee significante indirecte paden vanuit Intrinsieke waarde en Taakwaardering richting Autonomie, waardoor de twee positieve directe regressiepaden naar Autonomie vanuit Intrinsieke waarde en Zelfeffectiviteit nog eens extra wordt ondersteund.

Ook het regressiepadpatroon naar Afhankelijkheid is verschillend in beide groepen. Het pad van Testangst naar afhankelijkheid is uniek voor COMP en verklaart (daar) de meeste variantie in Afhankelijkheid (29% van de 36%). De attitudecomponenten verklaren in COMP nauwelijks iets in Afhankelijkheid (1%) en de overige 6% komt voor rekening van een cluster aan leerstrategieën dat bestaat uit Herhalen, Kritisch denken en Elaboreren. Ook het pad van Elaboreren naar Afhankelijkheid is uniek voor COMP. In VAK daarentegen wordt niet alleen veel minder variantie in Afhankelijkheid verklaard maar komt het grootste deel daarvan (7% van 11%) voor rekening van de attitudecomponent Cognitieve competentie. Herhalen en Kritisch denken nemen samen de resterende 4% verklaarde variantie voor hun rekening.

## 5.4 Conclusie en discussie

### 5.4.1 *Conclusie effecten van competentiegericht versus vakgericht statistiek-onderwijs*

De vraag in hoeverre de competentiegerichte leeromgeving aan de verwachting beantwoordt dat het – in vergelijking met de vakgerichte onderwijsbenadering – studenten tot een positievere attitude brengt en hen beter motiveert tot de inzet van diepgaander leerstrategieën en hogere leeruitkomsten realiseert, kan op basis van dit crossectionele onderzoek bevestigend worden beantwoord. Deze beantwoording moet als voorlopig worden opgevat, omdat slechts het allereerste deel van de nieuwe competentiegerichte leer- en toepassingslijn in het curriculum is onderzocht.

Op één uitzondering na blijken alle variabelen in het determinantenmodel gunstig uit te vallen voor het competentiegerichte ontwerp. Deze uitzondering betreft de variabele Herhalen, die doorgaans wordt beschouwd als een oppervlakkige leerstrategie. Het is verbonden met steeds terugkerende (routine)leertaken met als doel de beoogde cognitieve vaardigheden door oefening zo veel mogelijk te modelleren en te automatiseren. Herhalen beoogt ook opname van kennis in het geheugen door repetitie. Het is een leerstrategie die vooral geschikt is voor relatief eenvoudige taken, steeds terugkerende routinetaken en voor het oproepen van informatie uit het werkgeheugen. De strategie is minder adequaat voor onthouden in het lange termijn geheugen. De strategie heeft invloed op aandacht en codeerprocessen maar is niet effectief bij het tot stand brengen van verbanden tussen informatie of bij de integratie van informatie met voorkennis. Blom, Severiens, Broekkamp en Hoek (2005) interpreteren herhalen als oppervlakkig leren, hoewel dat ter discussie wordt gesteld door Biggs (2003), die stelt dat herhalen alleen oppervlakkig is wanneer het wordt toegepast in situaties waarin andere strategieën (zoals elaboreren en kritisch denken) adequater zijn. Een

voorbeeld van een positief effect van herhalen is de 'paradox of the Asian student' (Kember, Wong, & Leung, 1999). Aziatische studenten blijken veel meer uit het hoofd te leren en te herhalen dan hun westerse collegastudenten. Desondanks streven zij hun westerse collega's qua leerprestaties vaak voorbij. Kember et al. schrijven dit toe aan het gelijktijdig memoriseren en aandachtig streven naar begrijpen, wat leidt tot betekenisvol leren (Boulton-Lewis, Pillay, & Wilss, 2004). Anderson (2000) verklaart dit positieve effect van herhalen door het regelmatig praktiseren in een betekenisvolle context. Hij beschouwt herhalen toch wel als zinvol voor de eerste cognitieve leerfase. Omdat het onderwijs in de statistiek (data-analyse) aan het begin van de studie wordt gegeven en studenten voor het eerst kennismaken met dit vak, moeten de opvattingen over herhalen wellicht worden herzien en ook als een positieve leerstrategie worden beschouwd. Overigens pleit deze bevinding er wel voor bij een toekomstige meet-schaal onderscheid te maken in functioneel en dysfunctioneel gebruik van herhalen als leerstrategie.

Ten aanzien van de andere variabelen kan gesteld worden dat deze volgens de verwachting gunstiger voor COMP uitvallen. Opmerkelijk is daarin de gemiddelde score op de statistiekkennistest. Studenten in COMP ( $M = -2.21$  in logits) scoren hoger dan in VAK ( $M = -2.91$  in logits). Dit is opmerkelijk omdat de studenten in VAK qua studie-belastinguren tweemaal zoveel tijd aan statistiekonderwijs kunnen besteden (zie tabel 5.2). Dit zou een tamelijk grote prestatie van het competentiegericht ontwerp zijn, ware het niet dat beide gemiddelden ruim onder het gemiddelde 0 van de Raschschaal liggen (vergelijkbaar met niveau D respectievelijk C volgens de SOLO-taxonomie, vergelijk tabel 5.8). De interpretatie van de verschillen tussen de twee gemiddelden zou beter kunnen luiden: de COMP-studenten scoren minder negatief dan de VAK-studenten. Uitgaande van deze cijfers is het hogere doch nog steeds onvoldoende gemiddelde voor statistiekkennis in COMP toch nog minder bezwaarlijk te noemen dan in VAK, want in de competentiegerichte leer- en toepassingslijn krijgen studenten in de bachelorfase van hun opleiding nog diverse keren te maken met het uitvoeren van data-analyses in steeds complexer wordende onderzoeken. In het vakgerichte onderwijs echter wordt daaraan na de cursussen Statistiek die in de eerste studiejaar worden gegeven, tot aan de bachelorthesis geen of nauwelijks aandacht besteed. Wat betreft de onvoldoende scoring zou de geconstateerde moeilijkheidsgraad van de items in de statistiekkennistoets een verklaring kunnen zijn (zie de map van personen en items in bijlage III.2). Overigens is een lage score op statistiekkennis in retentieonderzoek een gebruikelijk beeld. Garfield en Ben-Zvi (2007) halen in hun overzichtstudie onderzoeken aan van Mathews en Clark (2003) en Clark et al. (2003) die in diverse steekproeven op diepgaande wijze nagingen wat studenten die een statistiekcursus zeer succesvol hadden afgesloten (grade 'A') zes weken na het tentamen zich nog herinnerden over de behandeling van de centrale limietstelling. De resultaten die deze 'topstudenten' lieten zien, noemen Garfield en Ben-Zvi 'alarming' aangezien deze studenten niet verder kwamen dan aan te geven hoe een gemiddelde berekend moet worden (Garfield & Ben-Zvi, 2007, p 383).

Overigens zijn de geconstateerde effectgrootten niet groot. De analyses met MANOVA wijzen uit dat het effect van de leeromgeving slechts 13% van de variantie in de veertien variabelen verklaart. Per variabele lopen de effectgrootten bij de significant bevonden verschillen uiteen 1.9% tot 4.1%.

Opmerkelijk zijn de causale patronen die zijn vastgesteld in COMP en VAK. Ook al moet deze als zeer tentatief worden beschouwd, de bevinding dat de werking van de variabelen attitude en motivatie in beide groepen verschilt, maakt het interessant daar verder onderzoek naar te doen. Beide variabelen zijn in de probleemanalyse in hoofdstuk 1 genoemd als mogelijke verklaringen voor de problemen in het statistiekonderwijs. Het lijkt erop dat in vakgericht onderwijs de (negatieve) attitude een pregnantere rol vervult en dat motivatie door de onderwijsopzet nauwelijks wordt opgewekt, maar dat in de competentiegerichte leeromgeving motivatie veel meer op de voorgrond komt en attitude – voorzover negatief – neutraliseert. Intrinsieke waarde geeft aan dat het uitvoeren van de taak op zichzelf plezier verschaft en positief affect oproept. Taakwaardering reflecteert het nut en het belang van de taak vanwege haar relatie met toekomstige of extrinsieke doelen. Beide motivatiecomponenten blijken in COMP unieke significante paden te vertonen die afwezig zijn in VAK, zoals het pad van Taakwaardering naar Kritisch denken en het pad van Intrinsieke waarde naar Autonomie. Beide motivatiecomponenten vertonen tevens significante positieve indirecte paden richting Autonomie.

Kritisch denken wordt in COMP ondersteund door drie (in plaats van twee, zoals in VAK) motivatiecomponenten. Kritisch denken wordt als één van de belangrijkste vormen van denken – hogere orde denken – voor het leren van statistiek beschouwd (Pfannkuch, 1999). Het feit dat het pad naar Kritisch denken wordt ondersteund door drie verschillende motivatiecomponenten, betekent dat er meer veerkracht is in geval van tegenvallers (het ‘multiple goal perspective’, zie bijvoorbeeld Lonka, Olkinuora, & Mäkinen, 2004), vooral ook omdat gebleken is dat Intrinsieke waarde, Taakwaardering en Zelfeffectiviteit in COMP significant hoger zijn dan in VAK. De bijdrage van Zelfeffectiviteit zorgt ervoor dat men verwacht dat men de taak succesvol zal kunnen invullen. Het feit dat Kritisch denken met deze drie motivatiecomponenten is verbonden, maakt dat Kritisch denken in COMP optimaal wordt ondersteund. Een negatieve relatie tussen Affect en Kritisch Denken zoals in VAK is afwezig in COMP. Het feit dat Intrinsieke waarde, Taakwaardering alsmede Zelfeffectiviteit in COMP hoger zijn dan in VAK en dat zowel Kritisch denken als Zelfregulatie in COMP uitsluitend door motivatiecomponenten in combinatie worden beïnvloed, geeft aanleiding om te concluderen dat Kritisch denken en Zelfregulatie als leerstrategieën optimaal worden ondersteund door de competentiegerichte leeromgeving.

Het pad vanaf Intrinsieke waarde naar Autonomie is eveneens significant en exclusief voor COMP. In COMP versterken Intrinsieke Waarde en Taakwaardering elkaar bij alle leerstrategieën (behalve bij Herhalen) en daaruit valt af te leiden dat het voldoende



hebben in de taak en het ervaren van het nut en het belang van de taak om een psychologisch probleem op te lossen (een extrinsiek doel), substantiële bijdragen leveren aan de beoogde zelfstandigheid in onderzoek doen, vooral omdat zij dit doen in combinatie met een in COMP uniek significant verband tussen Intrinsieke Waarde en Autonomie, ondersteund door twee voor COMP unieke significante indirecte paden naar Autonomie vanuit zowel Intrinsieke waarde en Taakwaardering. Hierdoor wordt Autonomie ondersteund en versterkt. De gemiddelde Autonomie in COMP is significant en substantieel hoger dan in VAK.

Een ander exclusief zichtbaar aspect van COMP, in vergelijking met VAK, is dat Testangst een significant verband laat zien met Afhankelijkheid. Afhankelijkheid wordt, in tegenstelling tot Autonomie, niet significant beïnvloed door de leeromgeving. Testangst echter is in COMP significant minder bepalend dan in VAK. De ondermijnende invloed van angst op het leren van statistiek en op de leerresultaten is uitgebreid bestudeerd en beschreven, hoewel meestal attitudes ten aanzien van statistiek worden bestudeerd (zie onder anderen Curda, 1997; Harlow, Burkholder, & Morrow, 2002; Sorge & Schau, 2002; Wisenbaker, Scott, & Nasser, 1999, 2002). Het onderzoek naar de angst voor statistiek en vervolgens het leggen van de focus op attitude ten aanzien van statistiek komt waarschijnlijk voort uit drie factoren: 1) het bij voorkeur toeschrijven van angst voor statistiek aan persoonseigenschappen (zie onder anderen Baluğlo, 2001) 2) de met de persoonlijkheid verweven, ingenestelde attitudes met hun 'resistance to change' (Gal, Ginsburg, & Schau, 1997) en 3) de onderzoekstraditie naar attitudes vanuit het wiskundedomein die sterk geworteld is in het behaviorisme en vanuit dat perspectief de student beschouwt als een stimulus-responsmachine. Dit wekt de indruk van een 'blame the student'-perspectief (Biggs, 2003). Het model toont aan dat zelfgereguleerde motivatie, die tot stand komt door een actieve interactie tussen leeromgeving en de student, positief bijdraagt aan het leerproces en de leerresultaten bij statistiek, wat betekent dat de leeromgeving zelf sterk bijdraagt aan deze interactie. Het toont ook aan dat het optimaliseren van de taak en het afstemmen van de leertaak op de doelgroep, uitermate positief bijdraagt aan de Taakwaardering en daarmee het mogelijk negatieve effect van testangst neutraliseert.

#### 5.4.2 *Beperkingen van het onderzoek*

Hoewel de bevindingen indicaties opleveren voor het positieve effect van de competentiegerichte leeromgeving op attitude, motivatie, leerstrategie en leeruitkomsten, is vanwege enkele onderzoeksbeperkingen enige terughoudendheid bij het trekken van conclusies gewenst. De onderzoeksbeperkingen betreffen de onderzoeksopzet (crossectioneel), de omvang van de onderzoeksgroepen, de nonrespons en het meetinstrumentarium.

Wat betreft het doen van uitspraken over de werking van het causale model is crossectioneel onderzoek niet voldoende. Zoals in paragraaf 5.2.4 al is aangestipt zijn volgens Cole en Maxwell (2003) voor mediatietoetsing minstens drie meetmomenten noodza-

kelijk zodat relaties vastgelegd kunnen worden tussen de onafhankelijke variabele(n) op tijdstip 1, de mediërende variabele(n) op tijdstip 2 en de uitkomstvariabele(n) op tijdstip 3. Wat betreft de toetsing van causale structuur van het theoretisch model met latente variabelen is ook de omvang van de onderzoeksgroep te klein.

De respons in dit onderzoek is tamelijk laag. In het algemeen gaat men ervan uit dat een hoge respons (70-80%) voldoende is om generaliserende uitspraken te doen. Door Ellemers (2004) is dit echter betiteld als het '*schandaal van de nonrespons*'. Hiermee doelt hij op de gangbare praktijk om te ontkennen of te bagatelliseren dat de uitkomsten van een surveyonderzoek ook bij een nonrespons van slechts 20% of 30% dubieus kunnen zijn. Deze visie van Ellemers wordt ondersteund door Stoop (2005) die in haar proefschrift concludeert dat niet de omvang van de groep nonrespondenten zorgt voor een vertekening van de resultaten, de zogenaamde nonresponsbias, maar de kenmerken van deze groep. Is er namelijk sprake van nonrespons MCAR (missing completely at random), dan leidt dit *niet* tot vertekening. Sterker nog, in een dergelijk geval kan een hogere respons zelfs leiden tot meer vertekening. De hoogte van de initiële response is dus volgens Stoop niet van belang als maat voor de mate van vertekening. Volgens Stoop (2005) is er een grotere kans op vertekening als de nonrespons samenhangt met het onderwerp van het onderzoek. Als dit gegeven wordt toegepast op dit onderzoek, dan betekent dit dat de kans op vertekening groter is als de reden voor niet-deelname vooral te maken heeft met het feit dat men een hekel heeft aan statistiek. Als men vooral situationele redenen heeft voor niet-deelname (geen tijd, nog niet aan cursus begonnen of gestopt met studie, te lange vragenlijst) dan is de kans op vertekening kleiner. Op basis van het vervolgonderzoek naar de nonrespons (zie bijlage III.1) kan worden geconcludeerd dat redenen voor niet-deelname vooral zijn gelegen in de situationele factoren. Hiermee zijn er geen aanwijzingen voor vertekening verkregen.

Voor het meten van de variabelen is gebruikt van een vragenlijst die aan studenten is voorgelegd. Dit houdt zelfrapportage van studenten in. Houkes (2002) brengt naar voren dat aan zelfrapportage het bezwaar kan kleven van verstoringe subjectieve invloeden van sociale en persoonlijke aard, alsmede van het probleem van gemeenschappelijke variantie van de methode ('common method variance'). Bij metingen met meer 'objectieve' variabelen zoals observatie en registratie van feitelijk gedrag zouden deze verstoringen vermeden kunnen worden. Daarentegen noemt Houkes (2002) ook enkele argumenten die juist wel pleiten voor het gebruik van de meer subjectieve metingen. Ten eerste verwijst ze naar een aantal studies waarin wordt aangetoond dat de correlaties tussen de 'objectieve' en 'subjectieve' metingen matig tot hoog zijn, waarbij ze verwijst naar Spector (1992) die op dit terrein onderzoek heeft verricht. Ten tweede wijst ze erop dat objectieve metingen ook hun problemen en beperkingen kennen: observaties door experts kunnen net als zelfrapportage van studenten worden beïnvloed door cognitieve en emotionele processen. Bovendien kunnen studenten zich ook anders gedragen dan gewoonlijk indien ze weten dat ze worden geobserveerd. Al met al kunnen er volgens Houkes ook vraagtekens worden geplaatst bij de

mate van objectiviteit van zogeheten objectieve metingen. Een derde argument om wel gebruik te maken van zelfrapportagevragenlijsten betreft de eenvoud van deze techniek: de toepassing van objectieve meetinstrumenten vergt meer inspanning en substantieel meer tijd en geld. Vragenlijsten daarentegen geven de mogelijkheid snel en voor relatief weinig geld grote hoeveelheden data te verzamelen. Resumerend kan worden gesteld dat er geen steekhoudende argumenten zijn tegen het gebruik van zelfrapportagevragenlijsten. Ten aanzien van de statistiekkennis is een met behulp van het Raschmodel gevalideerde, SOLO-geclassificeerde prestatiemeting uitgevoerd. Hiermee is de kans op het optreden van gemeenschappelijke-methodevariantie zo veel mogelijk gereduceerd.

De gebruikte meetinstrumenten voldoen aan zeer strenge Raschmodelcriteria. Hierdoor is eendimensionaliteit, betrouwbaarheid en validiteit (i.c. constructvaliditeit) van de meetinstrumenten beter gewaarborgd dan met Likertschaalprocedures (CFA) mogelijk zou zijn geweest. Dat neemt niet weg dat gebleken is dat de meetinstrumenten niet alle optimaal zijn. Vooral de moeilijkheidsgraad van sommige schalen en de spreiding van de items over het onderliggend continuüm laat in enkele gevallen te wensen over (zie de maps van personen en items in bijlage III.2). Voor verder onderzoek verdient het aanbeveling alle meetinstrumenten verder te systematiseren.

#### 5.4.3 *Slotbeschouwing*

Ondanks de onderzoeksbeperkingen die nopen tot een voorzichtige beoordeling van de onderzoekbevindingen, kan met vertrouwen gesteld worden dat de interventie met het competentiegericht statistiekonderwijs een succesvolle belofte inhoudt.

## **Van ontwerp naar implementatie: longitudinale bevindingen**

### **6.1 Inleiding**

In hoofdstuk 5 is uit de crosssectionele bevindingen gebleken dat studenten die het competentiegerichte statistiekonderwijs volgen zich meer betrokken voelen bij het leren en betere leerresultaten behalen dan degenen die studeren volgens de vakgerichte benadering. Deze bevindingen zijn gebaseerd op het eerste onderdeel van de onderzoekscompetentielijnen een jaar na start van de vakgerichte statistiekcursus (2004) of het onderzoekspracticum (2005). De vraag is of deze effecten ook op langere termijn standhouden. In dit hoofdstuk wordt een verkennend onderzoek hiernaar beschreven. Dit verkennend onderzoek is mogelijk doordat studenten die in 2002 een IMTO-pilot hebben gevolgd (zie hoofdstuk 4) of zich in datzelfde jaar voor een M&T-cursus hebben ingeschreven een vragenlijst hebben ingevuld die enkele meetinstrumenten gemeen heeft met de vragenlijst uit hoofdstuk 5 die deze studenten ook is voorgelegd. De meetinstrumenten die beide vragenlijsten gemeenschappelijk hebben, gaan over motivatie, leerstrategieën en statistiekkennis. Dit hoofdstuk beschrijft de bevindingen over de effecten van de leeromgeving op betrokkenheid bij het leren en leeruitkomsten drie jaar later in de tijd.

#### **6.1.1 Onderzoeksvraag**

In dit onderzoek worden twee leeromgevingen voor het statistiekonderwijs onderscheiden, te weten: de competentiegerichte en de vakgerichte leeromgeving. De verwachting bij de competentiegerichte leeromgeving is dat deze de student beter motiveert tot actieve en diepgaande bewerking van het leermateriaal wat leidt tot betere leeruitkomsten dan de vakgerichte leeromgeving, ook in de tijd. De vraag die in dit hoofdstuk centraal staat luidt dan ook:

*‘In welke mate is de competentiegerichte leeromgeving op langere termijn effectiever qua effect op motivatie, leerstrategieën en statistiekkennis dan de vakgerichte leeromgeving?’*

Er is naar verwachting sprake van interactie tussen leeromgeving en tijd, waarbij in de competentiegerichte leeromgeving (COMP) in vergelijking met de vakgerichte leeromgeving (VAK) motivatiecomponenten ten aanzien van onderzoek positiever zullen zijn en blijven en dat studenten adequatere leerstrategieën hanteren, wat resulteert in positievere leeruitkomsten.

## 6.2 Methoden

### 6.2.1 Respondenten

Van de 156 studenten waaraan voorafgaand aan de pilots met de competentiegerichtte leeromgeving (IMTO-5 óf IMTO-bètarun) een vragenlijst is afgenomen, hierna te noemen de  $T_1$ -vragenlijst (bijlage II.3), hebben er drie jaar later 25 de crosssectionele vragenlijst uit hoofdstuk 5, hierna te noemen de  $T_2$ -vragenlijst (bijlage II.2) ingevuld (respons: 16%). De  $T_2$ -vragenlijst is ook ingevuld door 168 van de 666 studenten (respons 25%) die de  $T_1$ -vragenlijst hebben ingevuld. De onderzoeksgroep bestaat dus uit 193 studenten.

Voor deelname aan een van de twee IMTO-pilots was succesvolle afsluiting van de cursus *Statistiek 1* als voorwaarde gesteld. Studenten die zich inschreven voor de cursus *M&T 1* (met het programma Onderzoeksmanager) was sterk aangeraden deze statistiekcursus vooraf bestudeerd te hebben. Op de  $T_2$ -meting is de gemiddelde leeftijd van de studenten uit de competentiegerichtte IMTO-pilots (kortweg: COMP) 41.23 jaar met een standaarddeviatie van 8.25 jaar, de gemiddelde leeftijd van de studenten uit de vakgerichte leeromgeving (kortweg: VAK) is 41.68 jaar met een standaarddeviatie van 9.50 jaar. In COMP is 27% man en 73% vrouw, in VAK is 31% man en 69% vrouw.

### 6.2.2 Meetinstrumenten

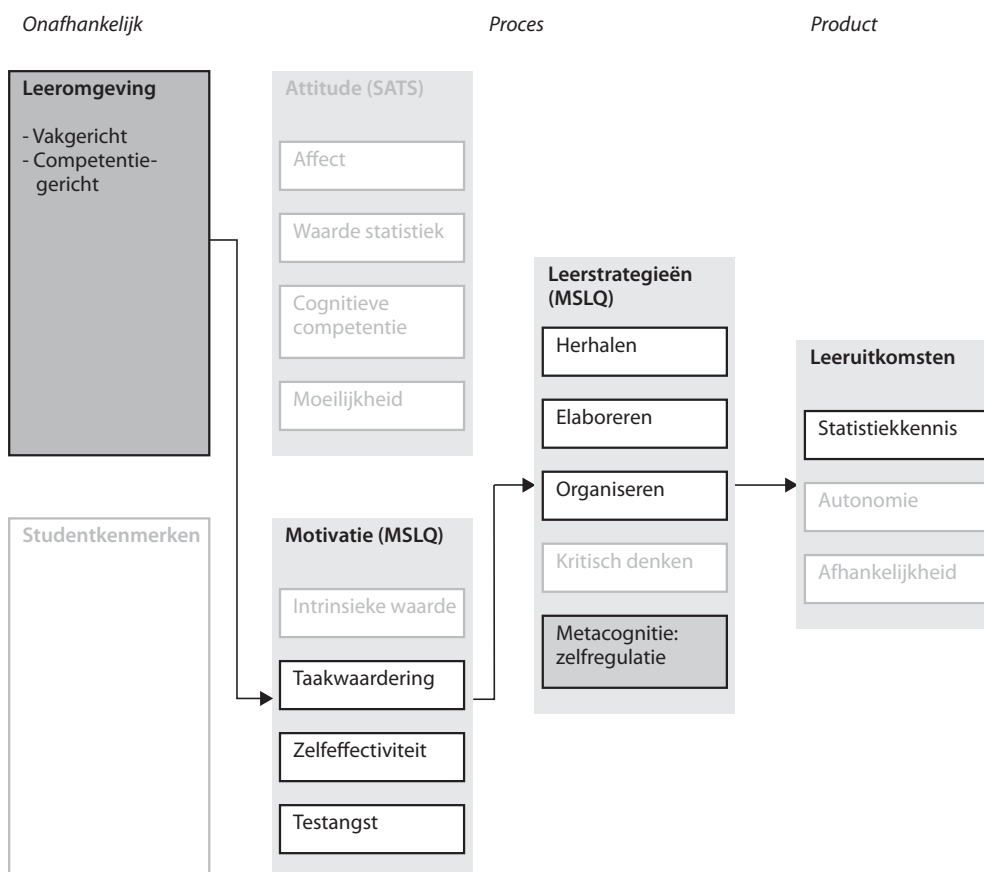
De  $T_1$ -vragenlijst (zie bijlage II.3) bevat voor motivatie en leerstrategieën minder items en schalen uit de MSLQ (Pintrich et al., 1991) dan de  $T_2$ -vragenlijst (bijlage II.2). De motivatie- en de leerstrategiecomponenten in de  $T_1$ -vragenlijst waren gebaseerd op een eerdere, minder items bevattende versie van de MSLQ, te weten die van Pintrich en De Groot (1990). In hun verkorte MSLQ-vragenlijst zijn de volgende psychologische constructen opgenomen die de betrokkenheid bij het leren representeren (zie voor een beschrijving van de constructen hoofdstuk 5, paragraaf 5.2): Taakwaardering, Zelfeffectiviteit, Testangst, Leerstrategieën (combinatie van Herhalen, Elaboreren en Organiseren) en Zelfregulatie. In deze lijst ontbreken de constructen Kritisch denken en Intrinsieke waarde. Ook de statistiekkennistest verschilt in beide vragenlijsten. Slechts drie van de op  $T_1$  gebruikte statistiekopgaven bleken voldoende psychometrische kwaliteit te hebben om te worden opgenomen in de  $T_2$ -kennistest. Uit de Rasch-analyses van de statistiekopgaven van de kennistest op  $T_2$  bleken echter van de drie oorspronkelijke items er twee niet te voldoen aan de criteria die het Raschmodel stelt (zie de voorstudie in bijlage III.2). Het enige resterende item waarop een vergelijking mogelijk is, betreft de vraag over de vuistregel van de standaardafwijking waarmee uitspraken kunnen worden gedaan over de spreiding van de scores rondom het gemiddelde. Tabel 6.1 bevat de items die zowel in de  $T_2$ -vragenlijst als in de  $T_1$ -vragenlijst voorkomen. De nummers verwijzen naar de desbetreffende vraag in de vragenlijst. Deze items zullen na de bewerking volgens het Rasch Rating Scale Model dienen als operationalisatie van de constructen uit het theoretisch model. Figuur 6.1 geeft het

determinantenmodel met de geoperationaliseerde variabelen weer. De variabelen die in vergelijking met de crosssectionele analyse niet kunnen worden opgenomen, zijn in de figuur in grijswaarden weergegeven. De causale pijlen zijn weggelaten.

TABEL 6.1 Overlappende items voor  $T_1$ - $T_2$ -vergelijking.

<i>Itemnr</i> <i>T<sub>2</sub></i>	<i>Itemnr</i> <i>T<sub>1</sub></i>	<i>Item</i>
		<b>Taakwaarde</b>
8	56	Ik denk dat er interessante dingen te leren zijn in deze modules.
9	46	Ik denk dat wat ik leer van statistiek door onderzoek te doen, ook kan gebruiken bij andere vakken.
10	43	Ik vind het belangrijk om zowel onderzoek te leren uitvoeren als statistiek toe te leren passen.
12	54	Ik denk dat wat ik allemaal door het uitvoeren van onderzoek leer, belangrijk is voor de rest van mijn studie.
13	44	Ik leer graag iets over onderzoeksmethoden en statistiek.
14	59	Begrip van statistische procedures is belangrijk voor mij.
		<b>Zelfeffectiviteit</b>
15	52	Ik denk dat hoge cijfers voor de eindopdrachten van deze modules haalbaar zijn.
19	50	Ik ben er zeker van dat ik de onderzoekstaken heel goed zal kunnen uitvoeren.
20	47	Ik verwacht dat ik deze modules met heel goede resultaten zal afronden.
21	2	Vergeleken met andere studenten die onderzoek doen en daarbij statistiek toepassen verwacht ik het beter te zullen doen.
22	45	Ik ben er zeker van dat ik deze modules goed zal kunnen invullen.
23	48	Ik denk dat ik een goede onderzoeker ben in vergelijking met andere studenten die onderzoek moeten doen.
26	57	Ik denk dat ik in staat ben om de theorie in deze modules te begrijpen en daardoor onderzoek uit te kunnen voeren.
		<b>Testangst</b>
27	60	Als ik mijn eindopdracht maak denk ik eraan hoe slecht ik het doe.
28	51	Ik heb een ongemakkelijk en ellendig gevoel als ik een eindopdracht moet maken.
29	3	Ik ben zo gespannen tijdens het maken van mijn eindopdracht, dat ik de dingen die ik geleerd heb niet meer op de juiste wijze kan passen.
30	58	Ik maak me zorgen over de eindopdrachten.
		<b>Leerstrategieën</b>
33	69	Om mezelf te helpen de theoretische informatie te onthouden lees ik mijn aantekeningen telkens over.
34	61	Als ik de theorie en onderzoekstaken bestudeer probeer ik dat met informatie uit andere bronnen te integreren.
35	79	Tijdens het bestuderen van de theorie probeer ik de dingen die ik lees, te verbinden met wat ik al weet.
37	75	Als ik een onderzoekstaak bestudeer, probeer ik dat met informatie uit verschillende bronnen te verbinden.
38	74	Bij het uitvoeren van nieuwe onderzoekstaken maak ik gebruik van wat ik heb geleerd in andere opdrachten.
47	62	Bij het bestuderen van het theoretische gedeelte probeer ik me te herinneren wat op de bijeenkomsten is gezegd, zodat ik de betekenis en achtergronden van het onderzoek zo goed mogelijk begrijp.
49	68	Als ik een eindopdracht voorbereid probeer ik mezelf zo veel mogelijk dingen uit het theoretisch gedeelte te herinneren.

FIGUUR 6.1 Aangepast determinantenmodel



### 6.2.3 Procedure

De studenten die zich hebben aangemeld voor deelname aan de pilots met de competentiegericht leeromgeving, IMTO-5 en IMTO-bètarun (zie hoofdstuk 4), hebben de  $T_1$ -vragenlijst ingevuld in november 2001 en januari 2002 voordat de pilots van start gingen. Aan de studenten die in de periode februari 2002-september 2002 de cursus *Methoden en technieken van onderzoek 1* hadden besteld, is met het oog op toekomstige vergelijkingen dezelfde vragenlijst toegestuurd, vergezeld van een voorgefrankeerde enveloppe voor retourzending. Ook aan deze studenten is verzocht de vragenlijst vóórafgaand aan het bestuderen van de cursus in te vullen en te retourneren.

De  $T_2$ -vragenlijst is in september 2005 per e-mail aangekondigd bij al degenen die in het verleden hadden deelgenomen aan de  $T_1$ -vragenlijstronde. Een week later werd de  $T_2$ -vragenlijst op het intranet van de Open Universiteit Nederland (*Studienet*) ge-



plaatst. Na een week ontvingen zij wederom per e-mail de officiële uitnodiging, met daarin vermeld het webadres voor het invullen van de vragenlijst. Na enkele weken werden alle studenten die nog niet hadden gereageerd telefonisch benaderd. Studenten waarvan geen e-mailadres bekend was, ontvingen een schriftelijke aankondiging en een week later een schriftelijke vragenlijst met gefrankeerde antwoordenveloppe. Ook zij werden na enkele weken gebeld indien zij nog niet hadden gereageerd.

#### 6.2.4 *Analyseschema*

Aangezien beide vragenlijsten een aantal items uit de diverse schalen gemeen hebben, kunnen met het Raschmodel vergelijkingen worden gemaakt tussen de resultaten op  $T_1$  en  $T_2$ . De items in de  $T_2$ -vragenlijst van het longitudinale bestand ( $N = 193$ ) zijn geijkt ('anchored') op de moeilijkheidsparameters van de items van het crosssectionele bestand ( $N = 468$ ). Vervolgens zijn de  $T_1$ -items geijkt op de moeilijkheidsparameters van de  $T_2$ -items. Voor de Raschanalyses is gebruik gemaakt van het programma Winsteps versie 3.6 (Linacre, 2005). De gevolgde procedure van de ijking van de items in het Raschmodel wordt besproken in bijlage III.4.

Voor de toetsing van de interactie tussen leeromgeving en tijdstip worden gemengde variantieanalyses met twee factoren (Engels: 'mixed between-within ANOVA', Tabachnick & Fidell, 2001) toegepast.

### 6.3 Resultaten

De verwachting is dat in de competentiegerichte leeromgeving er een grotere toename is in motivatie, er een gunstig effect is op leerstrategieën én er betere leerresultaten worden gerealiseerd dan in de vakgerichte leeromgeving.

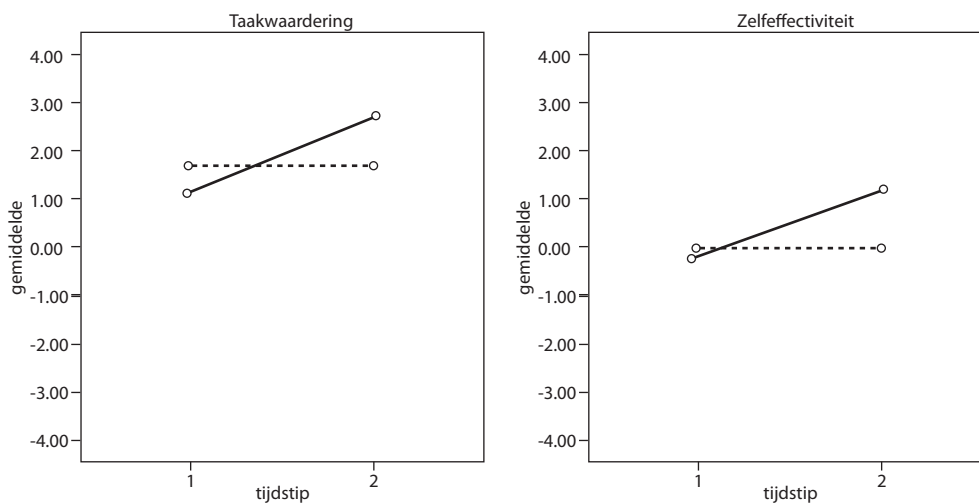
Volgens het determinantenmodel van figuur 6.1 betekent dit dat voor studenten die het competentiegerichte statistiekonderwijs (COMP) volgen de toename van de gemiddelde score op  $T_2$  voor Taakwaardering, Zelfeffectiviteit, Leerstrategieën, Zelfregulatie en Statistiekennis en de afname van Testangst hoger is in vergelijking met degenen die het vakgerichte statistiekonderwijs (VAK) volgen. Voor de toetsing van deze deelhypothesen met interactie-effecten zijn gemengde variantieanalyses uitgevoerd met leeromgeving en tijdstip als factoren. Aan de bij herhaalde metingen behorende assumptie betreffende de homogeniteit van intercorrelaties om deze variantieanalytische techniek te mogen toepassen, is bij alle verbanden voldaan. In tabel 6.2 zijn de resultaten van de interactietoetsing weergegeven. De gemiddelden en standaardafwijkingen zijn uitgedrukt in logits volgens het Raschmodel. In de praktijk variëren deze scores van - 4 (te interpreteren als 'gemakkelijk' of 'bijna iedereen mee eens') tot + 4 (te interpreteren als 'moeilijk' of 'bijna niemand mee eens').

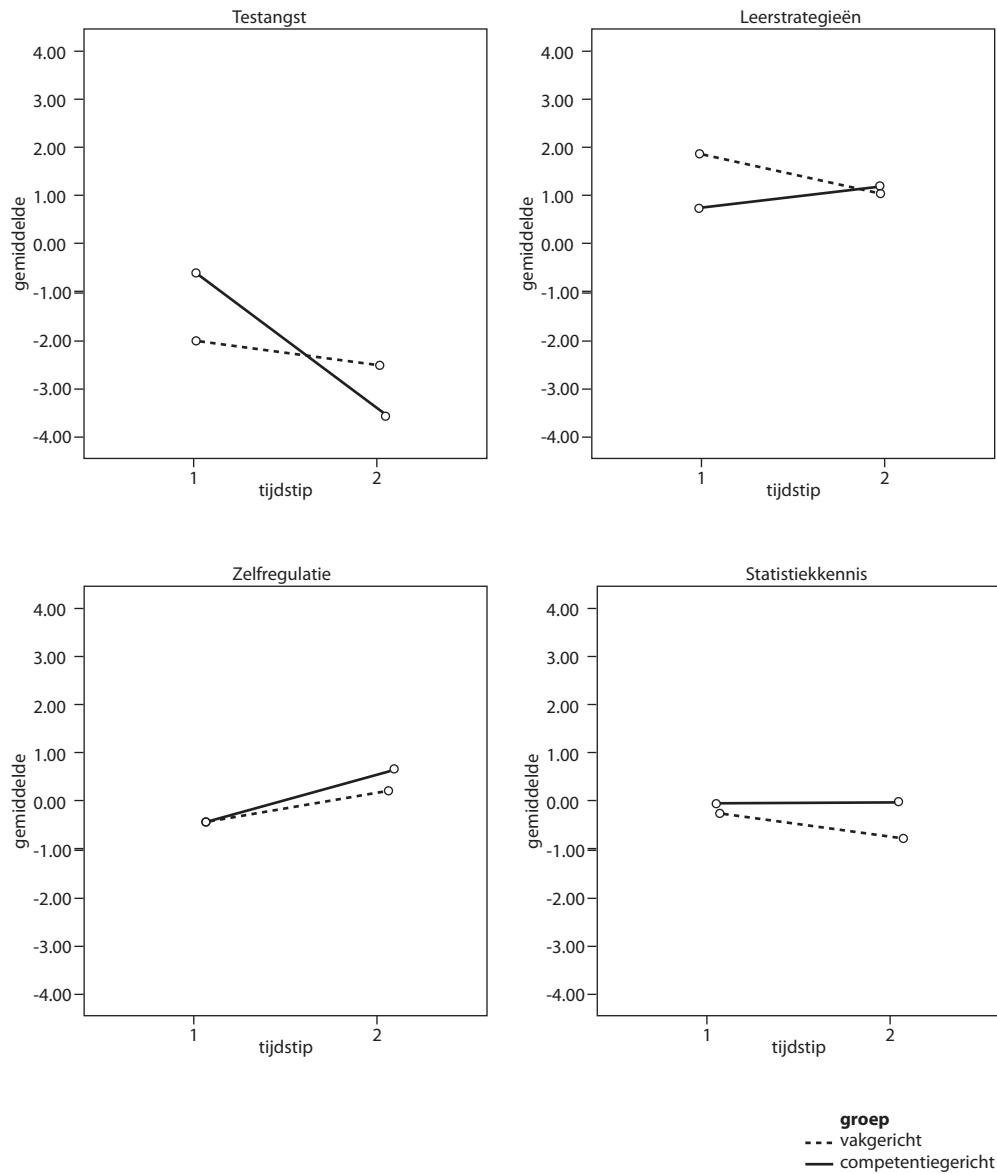
Alle zes interactie-effecten blijken, conform de verwachtingen, eenzijdige overschrijdingskansen te hebben die kleiner zijn dan  $2\alpha$  ( $\leq 10\%$ ) (Nijdam, 2004, p. 261). Twee daarvan – Zelfregulatie en Statistiekennis – blijken echter na Bonferroni-correctie niet statistisch significant. De interactie-effecten worden geïllustreerd in figuur 6.2.

De effectgrootten variëren volgens de indeling van Cohen (1988) van zwak tot matig ( $.01 < \eta^2 < .06$ ) voor de variabelen Taakwaardering, Zelfeffectiviteit, Zelfregulatie en Statistiekkennis en middelmatig ( $.06 < \eta^2 < .14$ ) voor de variabelen Testangst en Leerstrategieën.

TABEL 6.2 Resultaten van de interactie-effecten van leeromgeving en tijdstip

Construct	Tijdstip	COMP N = 25		VAK N = 168		Interactie-effect			Effectgrootte partiële $\eta^2$
		M	SD	M	SD	F	df	p	
Taakwaardering	$T_1$	1.11	1.79	1.63	2.65	9.004	1,191	.003	.045
	$T_2$	2.67	2.16	1.68	2.29				
Zelfeffectiviteit	$T_1$	-0.35	2.03	-0.20	2.47	7.043	1,191	.009	.036
	$T_2$	0.97	2.40	-0.13	2.56				
Testangst	$T_1$	-0.51	1.50	-1.98	2.44	14.402	1,191	.000	.070
	$T_2$	-3.38	3.12	-2.46	3.32				
Leerstrategieën	$T_1$	0.70	1.06	1.83	1.57	17.770	1,191	.000	.085
	$T_2$	1.14	0.86	1.00	1.21				
Zelfregulatie	$T_1$	-0.21	0.64	-0.16	1.32	3.661	1,191	.057	.019
	$T_2$	0.91	0.98	0.48	0.05				
Statistiekkennis	$T_1$	-0.08	0.97	-0.27	1.00	3.527	1,191	.062	.018
	$T_2$	-0.10	1.37	-0.83	1.25				





FIGUUR 6.2 Gemiddeldendiagrammen

## 6.4 Conclusie en discussie

### 6.4.1 *Conclusie effecten van competentiegericht versus vakgericht statistiekonderwijs*

De verwachting zoals die in paragraaf 6.2 is geformuleerd, is door de resultaten bevestigd: in COMP zijn de gemiddelden van de motivatiecomponenten Taakwaardering en Zelfeffectiviteiten significant hoger, van Testangst significant lager, wordt er significant diepgaander en gedisciplineerder gestudeerd (hogere gemiddelde waarden voor Leerstrategieën) dan in VAK. De effecten van Zelfregulatie en de retentie van de Statistiekennis zijn weliswaar conform het gestelde in de hypothese, maar bij toepassing van de Bonferroni-correctie statistisch niet significant. Het is de vraag of in dit stadium van het onderzoek (een eerste toetsing van een hypothetisch model) een conservatieve toetsing met strenge correctie niet te snel mogelijke voorspellers zou afschrijven (voor een discussie over het nut van de Bonferroni-correctie bij meerdere vergelijkingen op een databestand zie onder anderen Bland & Altman, 1995; Jones & Rushton, 1982; Perneger, 1995; Savitz & Olshan, 1995). Het verdient aanbeveling om in vervolgonderzoeken de rol van deze variabelen in het model te handhaven. De interactie-effecten duiden erop dat studenten die competentiegericht onderwijs volgen daarvan meer profijt lijken te hebben dan studenten die het vakgerichte onderwijs volgen. Dat uit de gemiddeldendiagrammen blijkt dat studenten in COMP bij de voormeting lager scoren op bijna alle variabelen (of hoger op Testangst) en dat daarna het beeld omslaat, is des te opmerkelijker. Dat COMP-studenten op  $T_1$  minder positief scoren dan de VAK-studenten zou erop kunnen wijzen dat eerdergenoemde studenten de IMTO-pilots hebben aangegrepen als een (laatste?) kans om alsnog de vakgerichte M&T- en statistiekcursussen succesvol af te sluiten en daarin zo goed zijn geslaagd dat de positieve werking voortduurt, ook na drie jaar.

Het hogere gemiddelde voor Taakwaardering op  $T_1$  in 2002 bij VAK zou kunnen betekenen dat vakgerichte studenten op dat meetmoment een groter belang toekenden aan het vak statistiek, gerelateerd aan bijvoorbeeld toekomstig beroep dan competentiegerichte studenten. Op  $T_2$  in 2005 is echter het gemiddelde voor Taakwaarde bij de vakgerichte studenten afgenomen en bij de competentiegerichte studenten toegenomen. Wellicht dat COMP-studenten in de loop van de tijd een groter belang zijn gaan hechten aan statistiek en de VAK-studenten juist minder het belang van dit vak gaan inzien voor bijvoorbeeld een toekomstig beroep. Dit resultaat zou mogelijk goed verklaard kunnen worden door het verschil in leeromgeving. De integratie van statistiek en methoden voor onderzoek en de inbedding van de leerstof in psychologische vraagstukken zorgt bij competentiegerichte studenten wellicht voor een groter begrip van het belang ervan en een hogere mate van waardering van het vak voor toekomstig functioneren.

VAK-studenten leken in 2002 minder angstig voor tests dan studenten uit de competentiegerichte leeromgeving. Het kan zijn dat met name studenten die angstiger

waren voor tests dan de gemiddelde student kozen voor deelname aan de nieuwe onderzoekspractica omdat ze er tegenop zagen de traditionele vakgerichte cursus te moeten doorlopen en af te ronden met een meerkeuzevraagtoets. In de loop van de tijd is bij de competentiegerichte studenten de angst voor tests gemiddeld meer afgenomen dan bij vakgerichte studenten en ook dit kan verklaard worden door het verschil in leeromgeving. Het cumulerende aspect in de nieuwe onderzoekspractica heeft als effect dat de student niet alleen in elk practicum nieuwe vaardigheden leert, maar tegelijkertijd ook genoodzaakt is om al verworven kennis en vaardigheden weer op te frissen en toe te passen. Met iets dat je regelmatig toepast, raakt men vertrouwd en boezemt daardoor in de loop van de tijd minder angst in.

VAK en COMP verschilden op  $T_1$  in 2002 niet in de mate waarin de studenten overtuigd waren van hun vaardigheden om succesvol te zijn in een cursus (Zelfeffectiviteit), maar de competentiegerichte studenten blijken in 2005 een grotere positieve verandering te vertonen wat Zelfeffectiviteit betreft dan de vakgerichte studenten. Een soortgelijke redenering als bij de verklaring voor het verschil in verandering in de mate van testangst geldt hier: van competentiegerichte studenten wordt bij elk volgend practicum gevraagd de verworven vaardigheden te integreren met de nieuw te verwerven kennis en vaardigheden. Daarbij is in de leerstof rekening gehouden met de betekenis, samenhang, timing, soort en hoeveelheid aan nieuwe informatie die studenten in hun leerproces kunnen verwerken. Dat kan het voor studenten in hun beleving gemakkelijker maken om een practicum succesvol af te ronden dan een traditioeel vakgerichte vak waarvan de leerstof minder nauw aansluit bij al aanwezige kennis. Het in eigen beleving op een constructieve manier werken aan een vak en vervolgens succesvol afronden kan als positieve bekrachtiging voor het eigen kunnen worden opgevat en dat vergroot de Zelfeffectiviteit (Bandura, 1986).

Opmerkelijk is het interactie-effect voor Leerstrategieën. In tegenstelling tot de verwachting dat er in VAK een geringere toename zou zijn in de tijd, blijkt bij Leerstrategieën er sprake te zijn van een significante *afname*. Er is sprake van een 'omgekeerd' effect in de betekenis dat VAK op  $T_1$  een gemiddeld hogere score blijkt te hebben dan COMP maar op  $T_2$  een significant lagere score dan COMP. Een mogelijke verklaring voor dit opmerkelijke verschijnsel kan wellicht gevonden worden in het commentaar van één van de respondenten in VAK die schreef: 'Ik heb het gevoel dat ik alleen maar kunstjes leer'. Een dergelijke perceptie van de gehanteerde strategieën lijkt een zwakke basis voor structureren en elaboreren en leidt wellicht tot een oppervlakkiger benadering van de stof.

Het moge duidelijk zijn dat dit onderzoek een sterk exploratief karakter heeft. Het signaleert enkele interessante resultaten die echter vooral duidelijk maken dat meer vervolgonderzoek noodzakelijk is om de verbanden tussen de variabelen nader te specificeren en te verklaren. In die zin kan het onderzoek waarover hier wordt gerapporteerd gekenschetst worden als een exploratief onderzoek dat bijdraagt aan het genereren van hypothesen voor vervolgstudies.

#### 6.4.2 *Beperkingen van het onderzoek*

Dit onderzoek heeft een aantal beperkingen die hieronder de revue passeren. De eerste beperking is dat er geen meting heeft plaatsgevonden op het moment dat de studenten hun cursus in VAK of COMP hebben afgerond. De reden daarvoor was dat de leerstof in beide leeromgevingen dermate verschillend zijn dat vergelijking van studenten op hun verworven statistiekkennis niet opportuun bleek te zijn.

Ook de meting van de statistiekkennis op de beide meetmomenten bleek niet probleemloos. Voor het onderzoeken van leeruitkomsten heeft dit onderzoek zich moeten beperken tot slechts één statistiekopgave die zowel op  $T_1$  en  $T_2$  aan de respondenten is voorgelegd. Dit item omvatte de vraagstelling 'Welke relatie bestaat er volgens u tussen de statistische grootheden 'gemiddelde', 'standaardafwijking' (ook wel standaarddeviatie genoemd) en 'normaalverdeling'? Beschrijf een bekende vuistregel voor deze begrippen in eigen woorden'. Het is in ieder geval bemoedigend om vast te stellen dat er voor het onthouden van zulke bekende statistische concepten als het gemiddelde, de standaarddeviatie en de normaalverdeling binnen COMP de retentie beter is dan in VAK (zie tevens de bevindingen die door Garfield en Ben-Zvi (2007, p. 383) over dit onderwerp zijn aangehaald). Herhalen behoort volgens Anderson (2000) nog tot de beginfase van het leren en geeft op zichzelf geen informatie over inzicht in verbanden met andere statistische concepten. Niettemin, het blijft een beperking van het huidige  $T_1$ - $T_2$  onderzoek dat leerresultaten niet onderzocht konden worden bij de moeilijkere opgaven uit de test die vooral betrekking hadden op begrip van variantieanalyse en regressie/correlatieanalyses.

Een andere belangrijke beperking is dat er in dit longitudinale onderzoek een betrekkelijk kleine groep (COMP,  $N = 25$ ) is vergeleken met een veel grotere groep (VAK,  $N = 168$ ). Ondanks deze beperking kan gesteld worden dat studenten die deelnamen aan COMP zich in ieder geval óók in de tijd significant en in positieve zin onderscheiden van de studenten die deelnamen aan VAK.

Een andere overweging die niet over het hoofd moet worden gezien is dat studenten van de Open Universiteit doorgaans langer over de studie doen dan reguliere studenten (gemiddeld drie modules per jaar). Over hun studieloopbanen is dan ook niets bekend. Deze situatie betekent tevens dat in zowel COMP als VAK meer verstorende invloeden een rol kunnen spelen die hun impact hebben op de hier gepresenteerde resultaten. Het lijkt echter onwaarschijnlijk om te veronderstellen dat die invloeden in COMP anders geweest zijn dan in VAK. De IMTO-pilots werden niet zelden gevolgd door studenten die dit als laatste mogelijkheid zagen om te slagen voor statistiek. De kleine COMP-groep in de longitudinale studie is eerder een representatie van dit type student dan van zelfgeselecteerde uitsluitend succesvolle studenten.

#### 6.4.3 Slotbeschouwing

Resumerend kan gesteld worden dat het evident is dat vervolgonderzoek met meer metingen in de tijd nodig is om uitsluitsel te geven over de resultaten en de factoren waaraan deze zijn toe te schrijven.





## Hoofdstuk 7

### Evaluatie: reflectie en discussie

#### 7.1 Inleiding

Dit proefschrift beschrijft het ontwerp van een interventie, de implementatie en de evaluatie ervan. De interventie betrof een competentiegerichte variant van het statistiekonderwijs in het psychologiecurriculum van de Open Universiteit Nederland. Paragraaf 7.2 bespreekt in het kort de hoofdstukken. Achtereenvolgens wordt aandacht besteed aan: de aanleiding voor de interventie, de leerpsychologische fundamenteën en de onderwijskundige inzichten die bij de ontwikkeling van de interventie zijn benut, het uiteindelijke ontwerp, de pilots waarin het ontwerp is getest en twee onderzoeken naar de effecten die met het ontwerp zijn gerealiseerd. Paragraaf 7.3 richt zich op de meest saillante opbrengsten van dit proefschrift en in paragraaf 7.4 komen de beperkingen van het proefschrift aan de orde. In paragraaf 7.5 volgen aanbevelingen voor follow-uponderzoek en paragraaf 7.6 bevat aanbevelingen voor de praktijk van het statistiekonderwijs.

#### 7.2 Samenvatting en beantwoording van de onderzoeksvragen

##### *Hoofdstuk 1*

Aanleiding voor dit proefschrift zijn de problemen die studenten hebben met het leren, herinneren en toepassen van statistiek. Hoofdstuk 1 geeft daarvoor drie verklaringen.

De eerste verklaring zoekt de oorzaak bij de studenten zelf: een deficiënte vooropleiding, een negatieve attitude, angst, geringe motivatie en weinig zelfvertrouwen. Veel onderzoeken op het terrein van het statistiekonderwijs laten een relatie zien tussen deze studentkenmerken en de leeruitkomsten. De tweede verklaring grijpt terug op de inhoud van het vak statistiek zelf. Deze is abstract en complex en wordt in compacte cursussen voornamelijk in de beginfase van de studie gegeven, op het moment dat studenten nog niet beseffen wat onderzoek inhoudt. In de literatuur ziet men allerlei pogingen om de statistiek voor studenten leerbaar en aangenaam te maken. Helaas blijken pogingen die tot nog toe binnen het curriculum zijn voorgesteld, niet structureel bij te dragen aan de oplossing van de hiervoor genoemde problemen die studenten ervaren (zie bijvoorbeeld Garfield & Ben-Zvi, 2007).

De derde verklaring zoekt de oorzaak in de positie die het vak statistiek in het curriculum heeft gekregen. Er wordt doorgaans een atomistische aanpak gehanteerd waarbij de leerstof gepresenteerd wordt in kleinere eenheden wat leidt tot fragmentatie van

het leerstofdomein en compartimentalisatie van kennis en vaardigheden (Van Merriënboer & Kanselaar, 2006, p. 279). Hierdoor komen de opbouw van het curriculum en daarmee het statistiekvak niet tegemoet aan de bijdrage die zij moeten leveren aan de verwerving van de onderzoekscompetentie.

Deze derde verklaring vormt het vertrekpunt van dit proefschrift. De interventie die in dit proefschrift centraal staat betreft het herontwerp van een gedeelte van het psychologiecurriculum, waarbij het conventionele statistiekonderwijs wordt vervangen door een competentiegerichte variant als mogelijke oplossing voor de problemen in het statistiekonderwijs.

### *Hoofdstuk 2*

De eerste onderzoeksvraag van deze studie luidt als volgt: *Welke determinanten bepalen het succes van het leren van statistiek?* Aangezien in dit proefschrift de verklaring van de problemen in het statistiekonderwijs wordt gelegd bij de positie die het vak in het psychologiecurriculum inneemt, staat in hoofdstuk 2 de vraag centraal welke determinanten in de onderwijsleeromgeving bepalend zijn voor het verwerven van de onderzoekscompetentie. Hoewel de beoogde onderzoekscompetentie een breed terrein aan leerinhouden (psychologie, M&T en statistiek) beslaat, wordt in dit proefschrift de aandacht specifiek gericht op de verwerving van de statistiekkennis die voor psychologisch onderzoek noodzakelijk is.

Bij de inventarisatie van de determinanten is voornamelijk aangesloten bij de theorie van zelfgereguleerd leren (Pintrich, 2004; Zusho & Pintrich 2003) die een onderscheid maakt in drie soorten variabelen: achtergrondvariabelen als studentkenmerken en kenmerken van de leeromgeving; procesvariabelen als motivatie en inzet van leerstrategieën en leeruitkomsten als productvariabelen.

Voorts is uitgegaan van de systematiek van Intervention Mapping (Bartholomew et al., 2006). Bij de afstemming van gedragsdoelen, determinanten en veranderdoelen zorgt Intervention Mapping er voor dat de samenhang daartussen zichtbaar wordt én blijft gedurende het ontwerpproces. Gedragsdoelen hebben betrekking op de gewenste leeruitkomsten. Veranderdoelen zijn doelen die specifiek aangeven wat er in de determinanten moet veranderen om de gewenste leeruitkomsten te verkrijgen. Door middel van uitgebreide matrices worden de relaties tussen de diverse doelen en determinanten gedetailleerd beschreven.

De belangrijkste en voor een onderwijsontwerper ook veranderbare determinanten om studenten op effectieve, doelmatige en plezierige wijze tot de beoogde leeruitkomsten te brengen, zijn gelegen in de leeromgeving en de daarbinnen gehanteerde instructiemethoden. Door een wijziging in de positionering van het vak statistiek waardoor het met M&T en psychologie kan worden geïntegreerd tot de onderzoekscompetentie, kan de synergie tussen inhoud, didactiek en technologie geoptimaliseerd worden. Dit impliceert onder andere de verdeling van de statistiekinhoud over meer cursussen

(van eenvoudig naar complex), het leren kennen van de functie, waarde en het nut van statistiek in betekenisrijke contexten door onderzoek uit te voeren, enzovoorts. Afstemming van deze determinanten op de gewenste leeruitkomsten of gedragsdoelen is afgebeeld in een matrix van veranderdoelen (bijlage I.1). Deze veranderdoelen vormen het concrete antwoord op de onderzoeksvraag wat in de leeromgeving moet veranderen om de gedragsdoelen te kunnen realiseren.

Naast de belangrijke determinanten voor het leren zijn ook variabelen als attitude, angst, motivatie en de studiestrategieën die studenten hanteren cruciale condities voor het leren. Deze persoonlijke determinanten bepalen als procesvariabelen de leerprocessen die zich afspelen in de relatie tussen de onderwijsleeromgeving en de leeruitkomsten. Ze worden in dit proefschrift echter niet beschouwd als direct veranderbare determinanten, maar als indirect veranderbare determinanten die via ingrepen in de leeromgeving kunnen worden beïnvloed. Een onderwijsleeromgeving die door studenten als interessant, uitdagend, niet bedreigend en plezierig wordt ervaren, verhoogt de motivatie om diepgaand te studeren en levert een bijdrage aan een positieve attitude ten aanzien van het vak.

Het inlassen van de procesvariabelen in het model leidt ertoe dat met Intervention Mapping op systematische wijze gespecificeerde beslissingen kunnen worden genomen over de veranderingen in de onderwijsleeromgeving die nodig zijn om óók de motivatie voor het doen van onderzoek, de attitude ten aanzien van statistiek te verbeteren en de inzet van adequatere leerstrategieën te bevorderen zodat mede als gevolg daarvan de kans op goede leeruitkomsten wordt geoptimaliseerd.

De inventarisatie van determinanten maakte inzichtelijk dat het thans beschikbare onderzoek naar de verwerving van complexe leerinhouden, zoals statistiek, niet altijd heldere en/of volledige antwoorden biedt voor het ontwerp van een leeromgeving.

### Hoofdstuk 3

In hoofdstuk 3 staat de vraag centraal: *Welke onderwijskundige instructieaanwijzingen zijn effectief om de gewenste veranderingen in de leeromgeving te bewerkstelligen die leiden tot succes van het leren van statistiek?* Ter beantwoording van die onderzoeksvraag zijn methodieken en praktijken uit de 'Cognitive Load Theory' (Van Merriënboer & Sweller, 2005), 'Meaningful Learning' (Ausubel, 1960, 1968), het 'Cognitive Apprenticeship Model' (Collins, Brown, & Newman, 1989) en de Theory of the Research Environment (Gelso, 2006) geselecteerd die aansluiten bij de wijze hoe mensen complexe en abstracte kennis verwerven en die toepasbaar zijn in het domein van het uitvoeren van psychologisch onderzoek. De meest belangrijke instructieaanwijzingen zijn: het activeren van voorkennis, het gebruik van advance organizers, gelegenheid bieden tot oefenen en toepassen, het voorkomen van cognitieve overbelasting en het gebruik van een variëteit van authentieke onderzoeksvraagstukken.

#### Hoofdstuk 4

In hoofdstuk 4 staat de derde vraag centraal: *Hoe kunnen de instructieaanwijzingen worden gesynthetiseerd tot een ontwerp voor een effectieve leeromgeving dat kan worden geïmplementeerd in een bestaand, vakgericht curriculum?* Het geheel aan instructieaanwijzingen en de simultane toepassing daarvan sluit nauw aan bij het Four-Component Instructional Design Model (4C/ID-model) van Van Merriënboer (1997). Dit model sluit met zijn systematische aanpak van realistische, authentieke 'hele-taken' als 'motor' voor het leren naadloos aan op het doen van onderzoek om al onderzoekende zich de competentie en daarmee de statistische kennis eigen te maken. Om het leren met complexe, realistische taken in goede banen te leiden, wordt er voorzien in relevante informatie, feedback, uitgewerkte voorbeelden, begeleiding en scaffolding. In bijlage I.4 is de blauwdruk weergegeven van de onderzoekscompetentie volgens het 4C/ID-stramien. Hierin wordt onder meer gerepresenteerd hoe de statistiekleerstof in een lange leer- en toepassingslijn over de gehele bachelorperiode is verdeeld. Het 4C/ID-model is op een pragmatische wijze gehanteerd om de verzamelde instructieaanwijzingen daarin onder te brengen. De gehanteerde werkwijze van de ontwerper blijkt sterk overeen te komen met het hedendaagse, meer praktijkgerichte gebruik van het model, zoals dat door Van Merriënboer en Kirschner (2007) tien jaar na de introductie ervan is omschreven.

Het ontwerp van de onderzoekscompetentie is in pilots met studenten van de opleiding Psychologie van de Open Universiteit Nederland in een elektronische leeromgeving beproefd (IMTO-project, Van Buuren & Giesbertz, 1998; Giesbertz & van Buuren, 2004). Deze pilots hadden als doel na te gaan of studeren en werken aan onderzoek in een elektronische leeromgeving mogelijk en werkbaar is en in kaart te brengen of studenten tevreden zijn en de nieuwe aanpak waarderen. Gebleken is dat het aanbieden van het onderwijs in de elektronische leeromgeving niet tot problemen hoeft te leiden. Ook stellen studenten zich positiever op ten opzichte van de statistiek en lijken ze gemotiveerder om de leerstof diepgaand te bestuderen. Wat niet uit de pilots kan worden opgemaakt, is of studenten ook betere leerresultaten behalen dan in de conventionele aanpak van het statistiekonderwijs. Toch waren de pilots veelbelovend genoeg om door te gaan en besloten werd het ontwerp in de vorm van onderzoekspractica in het curriculum van de opleiding Psychologie in te passen.

De onderzoekspractica zijn geordend van eenvoudig naar complex, zijn cumulatief en dienen in een bepaalde volgorde te worden doorlopen. Duidelijk is geworden dat er voor de implementatie van een dergelijk geïntegreerd ontwerp aan verschillende organisatorische randvoorwaarden moet worden voldaan, zoals teamwerk tussen inhoudsdeskundigen uit de drie verschillende disciplines (psychologie, M&T en statistiek), bewaking van de competentielijn gedurende de gehele bachelorperiode, bewaking van het competentiegericht concept, vaardigheid in e-didactiek, ingangseisen voor deelname aan een onderzoekspracticum en een goed functionerende elektronische leeromgeving.

Het ontwerp is in de vorm van een reeks onderzoekspractica geïmplementeerd in het curriculum van de opleiding Psychologie. De verantwoordelijkheid voor elk practicum ligt bij een team van psychologie-, M&T- en statistiekdocenten. Door de integratie met vier andere cursussen kan de statistiek die voorheen in twee cursussen van elk 120 studiebelastinguren wordt bestudeerd, nu worden verspreid over een reeks van practica die in totaal 1200 studiebelastinguren beslaat (zes onderzoekspractica van elk 120 uur en de bachelorthesis van 480 uur). In twee onderzoekspractica ligt de nadruk op nieuwe statistiekleerstof.

#### *Hoofdstuk 5 en 6*

Om de vierde vraag te beantwoorden: *In welke mate is de competentiegerichte leeromgeving effectiever dan het vakgerichte curriculum in het succesvol leren van statistiek?* zijn twee empirische onderzoeken uitgevoerd onder de studenten van de opleiding Psychologie van de Open Universiteit Nederland, een crosssectioneel (hoofdstuk 5) en een longitudinaal (hoofdstuk 6) vragenlijstonderzoek. Beide onderzoeken hadden betrekking op het eerste herontworpen practicum *Kwantitatieve data-analyse* dat wordt vergeleken met de statistiekcursussen.

Het crosssectionele onderzoek is gehouden onder studenten die de vakgerichte cursussen statistiek hebben gevolgd (VAK,  $N = 340$ ) en studenten die het nieuwe, competentiegerichte practicum hebben gevolgd (COMP,  $N = 128$ ). De vraag in hoeverre het competentiegerichte onderwijs aan de verwachting beantwoordt dat het – in vergelijking met de vakgerichte onderwijsbenadering – studenten tot een positievere attitude brengt en hen beter motiveert tot de inzet van diepgaander leerstrategieën en hogere leeruitkomsten realiseert, kan op basis van het crosssectionele onderzoek bevestigend worden beantwoord. De scores op de attitude-, motivatie-, leerstrategieën-, autonomie- en de afhankelijkheidsschalen vallen voor COMP gunstiger uit dan voor VAK. Ook scoort COMP hoger op de statistiekkennistest. Dit is opmerkelijk omdat de studenten in VAK qua studiebelastinguren tweemaal zoveel tijd aan statistiekonderwijs kunnen besteden (zie tabel 5.2). Overigens was in beide gevallen het niveau van de statistiekkennis onvoldoende. De interpretatie van de verschillen tussen de twee gemiddelden zou beter kunnen luiden: de COMP-studenten scoren minder negatief dan de VAK-studenten.

Uitgaande van deze cijfers is het hogere doch nog steeds onvoldoende gemiddelde voor statistiekkennis in COMP toch nog minder bezwaarlijk te noemen dan in VAK, want in de competentiegerichte leer- en toepassingslijn krijgen studenten in de bachelorfase van hun opleiding nog diverse keren te maken met uit te voeren data-analyses in steeds complexer wordende onderzoeken. In het vakgerichte ontwerp daarentegen wordt daaraan na de cursussen Statistiek die in de eerste studiejaar worden gegeven, tot aan de bachelorthesis geen of nauwelijks aandacht besteed.

Een andere interessante bevinding betreft de verschillende rollen van attitude en motivatie. Hoewel de modeltoetsing in hoofdstuk vijf met behulp van Structural Equation Modeling tentatief was, mag voorzichtig geponeerd worden dat het volgende

patroon zich voordoet: bij VAK domineert attitude en komt motivatie niet of nauwelijks in beeld, bij COMP lijkt attitude te worden geneutraliseerd en speelt motivatie een grotere rol. Dit zou een mogelijke verklaring kunnen zijn van de problemen die studenten ervaren in het statistiekonderwijs (zie hoofdstuk 1).

In hoeverre de resultaten ook na verloop van tijd gunstig blijven uitvallen voor COMP, is nagegaan in het longitudinale onderzoek. Daartoe is de vragenlijst uit het crosssectionele onderzoek in september 2005 ook voorgelegd aan studenten die in de eerste helft van 2002 aan een van de twee pilots (IMTO-bètarun of IMTO-5) hebben deelgenomen ( $N = 156$ ) of die zich in die periode voor een cursus M&T hebben ingeschreven ( $N = 666$ ). Ook na drie jaar blijkt COMP hoger te scoren dan VAK op de motivatiecomponenten en de inzet van meer diepgaande leerstrategieën.

Hoewel de uitkomsten gunstig uitvallen voor COMP zijn de effectgrootten niet groot. Proporties verklaarde variantie in de statistiekkennis van 8% in COMP en 19% in VAK met twaalf factoren zijn niet overtuigend. Deze cijfers zijn overigens niet veel lager dan in ander onderwijskundig onderzoek.

#### *Conclusie*

De onderzoeksresultaten laten zien dat de uitkomsten van het competentiegericht statistiekonderwijs in vergelijking met het vakgerichte onderwijs over de gehele linie positiever zijn en dat deze uitkomsten ook duurzaam zijn. Het competentiegericht statistiekonderwijs genereert goede resultaten en lijkt een oplossing te bieden voor de problemen in het statistiekonderwijs. Niet alleen wat betreft het leren, herinneren en toepassen van de statistiek, maar ook met betrekking tot de in hoofdstuk 1 genoemde problemen ten aanzien van attitude, angst, motivatie en oppervlakkige leerstrategieën. De effecten zijn echter niet groot. Hierin wijkt dit onderzoek niet af van wat elders in onderwijskundig onderzoek wordt gevonden.

### **7.3 Opbrengsten van het onderzoek**

In deze paragraaf worden de belangrijkste opbrengsten van het onderzoek besproken. Achtereenvolgens wordt aandacht besteed aan het herontwerp van het statistiecurriculum, de meetinstrumenten, de methodiek en evaluatie van de eerste resultaten en het conceptuele model.

#### *Het herontwerp*

Met het herontwerp heeft statistiek een geïntegreerde positie gekregen in het psychologiecurriculum. Statistiek, M&T en psychologie zijn nu horizontaal met elkaar verbonden in hele-onderzoekstaken. Dit leidt ertoe dat ook inhouds- en M&T-docenten meer betrokken en meer verantwoordelijk zijn voor de ontwikkeling en toepassing van statistiekkennis bij studenten. De langere leer- en toepassingslijn maakt het mogelijk dat de student van meet af aan betrokken is bij het uitvoeren van authentieke, aan



de psychologische praktijk ontleende onderzoekstaken, waar statistiek steeds weer wordt toegepast om antwoorden te formuleren op onderzoeksvragen.

In het herontwerp is in de inventarisatie van instructiemaatregelen uitdrukkelijk rekening gehouden met het te verwachten effect dat deze maatregelen kunnen hebben op de attitude ten aanzien van statistiek, motivatie om psychologisch onderzoek uit te voeren en de beoogde cognitieve en metacognitieve strategieën die gehanteerd worden door de student. De eerste resultaten van het herontworpen statistiekonderwijs lijken erop te wijzen dat de veranderingen in de leeromgeving haar vruchten heeft afgeworpen.

#### *Meetinstrumenten*

Voor het evalueren van het ontwerp is veel aandacht besteed aan de kwaliteit van de gehanteerde meetinstrumenten. Alle gebruikte instrumenten zijn gevalideerd, zowel in CFA (Arbuckle, 2003; Byrne, 2001) als in het Raschmodel (Linacre, 2005). Dat heeft tot gevolg gehad dat veel van de gebruikte schalen zijn gemodificeerd, aangepast of herontworpen, voordat de evaluatie van het ontwerp heeft plaatsgevonden. Voorafgaand aan de statistische analyses zijn alle gebruikte schalen in het Raschmodel getransformeerd tot intervalschalen, waardoor meetruis geen verstorende impact heeft op de gebruikte statistische technieken (Bond & Fox, 2001). Het determinantenmodel zelf is, gebruikmakend van de Raschschalen, gekruisvalideerd in SEM (Byrne, 2001). De meetinstrumenten kunnen in vervolgonderzoek verder worden verfijnd.

#### *Het gehanteerde model*

Het uitgangspunt was een aan Pintrich (2004) ontleend theoretisch model dat een aantal onderzoeksvariabelen in samenhang onderzoekt. Dit in tegenstelling tot veel onderzoeken waarin steeds sprake is van één van die variabelen in relatie tot de prestaties op het terrein van de statistiekkennis (Garfield & Ben-Zvi, 2007). Het model geeft daarmee aan dat er in de relatie tussen leeromgeving en leeruitkomsten mediërende processen plaatsvinden die verklaren waarom de ene onderwijsbenadering studenten beter weet te motiveren tot diepgaander bestudering van de leerstof dan de andere.

Ook lijkt het, vanuit een interventieperspectief, belangrijk om aandacht te besteden aan motivatie én attitude. Gebleken is dat een negatieve attitude, die tamelijk hardnekkig lijkt te zijn in veel onderzoeken, evenals angst voor de statistiek, niet persistent hoeft te zijn. In de toetsing van het determinantenmodel met Structural Equation Modeling (bijlage III.3) is bijvoorbeeld naar voren gekomen – zij het tentatief – dat attitude en motivatie verschillende uitwerkingen hebben in beide leeromgevingen. Dat in veel onderzoek (een negatieve) attitude ten aanzien van en angst voor de statistiek als een bijna onveranderlijke predictor voor de leeruitkomsten naar voren komt (Garfield & Ben-Zvi, 2007), heeft ook te maken met het feit dat dit soort onderzoeken altijd plaatsvinden binnen eenzelfde soort leeromgeving (geen controlegroep) en met autonoom statistiekonderwijs dat min of meer losstaand van het psychologiedomein.

*Intervention Mapping*

Voor de reconstructie van het ontwerpproces is gebruik gemaakt van Intervention Mapping. Dit protocol voorziet in een planmatige aanpak van beoogde gedragsveranderingen. Met behulp van transformatiematrices kan het ontwerpproces op transparante wijze gevolgd worden. Het gebruik van Intervention Mapping voor onderwijskundige interventies toont beperkingen die inherent zijn aan het domein en niet toe te schrijven zijn aan Intervention Mapping als zodanig. Door van Intervention Mapping gebruik te maken blijkt dat er nauwelijks tegemoet gekomen kan worden aan het uitputtend en in meetbare termen beschrijven van gedragsdoelen. Door gedragsdoelen volgens Intervention Mapping te formuleren blijkt dat er veelal geen één-op-eenrelatie bestaat tussen de specificaties en de wijze waarop de variabelen die de gedragsuitkomsten meten zijn geoperationaliseerd. Intervention Mapping schrijft ook voor determinanten die de gewenste gedragingen beïnvloeden idealiter te baseren op evidence-based theorieën, maar de op basis van gedragsdoelen en determinanten geformuleerde veranderdoelen blijken in de praktijk niet altijd gebaseerd te zijn op empirisch bewijs. Ook de koppeling van veranderdoelen aan instructieaanwijzingen uit de onderwijskundige literatuur kan vanwege de beperkte empirische evidentie niet geheel adequaat worden gefundeerd. Dit heeft onder andere gevolgen voor de mogelijkheden leerprocessen te optimaliseren. Toepassing van Intervention Mapping laat zien waar de ‘zwakke’ plekken zitten. Het wordt met name toegepast in de gezondheidswetenschappen voor relatief kleine en overzichtelijke interventies en is in de onderwijskunde nagenoeg onbekend.. De toepasbaarheid van Intervention Mapping is door vage, slecht afgebakende concepten (‘fuzzy concepts’: Van der Klink & Boon, 2003), gebrekkige empirische evidentie en een hoge mate van retoriek die veel onderwijskundige literatuur kenmerkt, dus lastig maar wel aanbevelenswaardig.

#### 7.4 **Beperkingen van het onderzoek**

*Generaliseerbaarheid van de resultaten*

De resultaten van het onderzoek zijn niet zonder meer generaliseerbaar naar andere universitaire psychologieopleidingen. Ten eerste is de doelgroep van studenten bij de Open Universiteit Nederland anders dan die van de reguliere universiteiten. De gemiddelde leeftijd van de studenten bij de Open Universiteit Nederland is hoger. Er is geen selectie aan de poort. De studenten combineren hun studie merendeels met een baan. Meestal studeren zij in deeltijd en verdelen ze hun tijd en energie over studie, baan, gezin en/of kinderen. Studeren bij de Open Universiteit Nederland is grotendeels gebaseerd op zelfstudie en op materialen en een onderwijsomgeving die deze studievorm optimaal ondersteunen. Dit kan betekenen dat de studenten meer werk- en/of levenservaring hebben, meer gewend zijn keuzen te maken en autonomer hun studie beoordelen. Daardoor zijn ze wellicht gemakkelijker in staat zijn om in een dergelijke lange leer- en toepassingsrij zelfstandig psychologisch onderzoek uit te voeren. Daar staat echter tegenover dat de Open Universiteitsstudenten, in tegenstelling tot studerenden aan de reguliere universiteiten, veel vaker geconfronteerd worden met werk-thuis-(studie)interferentie en/of (studie)-thuis-werkinterferentie

(Schaufeli, Bakker, & De Jonge, 2003). Deze overwegingen onderstrepen dat de bevindingen met het huidige statistiekerontwerp niet zonder meer generaliseerbaar zijn naar andere studentpopulaties.

#### *Organisatorische randvoorwaarden*

Het competentiegericht onderwijsontwerp gaat uit van horizontale integratie van de verschillende vakken die deel uitmaken van het onderzoekscurriculum (statistiek, M&T en psychologie). Dit betekent dat betrokken inhoudsdeskundigen nu ook de mogelijkheid en taak hebben de statistiek (en M&T, kortweg onderzoek) in hun vak een duidelijke plaats te geven en zo de professionele competenties van de studenten te bevorderen. Bovendien geeft het de inhoudsdocenten de mogelijkheid meer teamgericht te werken. Niet zelden blijkt dit aspect in de onderwijspraktijk één van de grootste struikelblokken voor het herontwerp van onderwijs. Het vergt een teaminspanning van vakspecialisten die onderwijskundige kennis en inzichten aanwenden ter wille van verwachtingen op het gebied van de beoogde effectiviteit van het geboden onderwijs. Meer nog, het vergt het loslaten en overstijgen van louter domeinspecifieke doelen ter wille van een meer diepgaande, geïntegreerde leeropbrengst. Dezelfde voordelen gelden voor statistiekdocenten (verantwoordelijk voor inhoud en meer teamgericht werken). Het behoeft geen betoog dat lokale belangen van verschillende disciplinair opgebouwde groepen (voorheen vakgroepen) hiermee in het geding komen. Alle betrokken specialisten worden uitgedaagd tot een werkwijze die beoogt aan te sluiten bij een meer generalistische benadering en die een groot beroep doet op het projectmatig samenwerken als team om tot inhoudelijke integratie te komen. Bestaande vakstructuren in organisaties kunnen een dergelijke teaminspanning belemmeren. Voor de opleiding Psychologie van de Open Universiteit Nederland hebben deze belemmeringen niet gespeeld. De faculteit bestaat uit een kleine staf. Zij heeft een onderwijsopvatting die voortvloeit uit de innovatiemissie die door het ministerie van Onderwijs bij de Open Universiteit Nederland is neergelegd. De faculteit kent een centrale aansturing bij het aanbieden van een beperkt aantal cursussen. Dat noodzaakt de staf tot samenwerken om doublures in het aangeboden materiaal te voorkomen. Bij andere universiteiten met grotere faculteiten zijn de organisatorische randvoorwaarden waarschijnlijk minder gunstig. Daar is veelal sprake van een aansturing die verdeeld is over verschillende disciplines met veel cursussen die zelfstandig en onafhankelijk van elkaar worden ontwikkeld en waar geen sprake is van een expliciete onderwijsinnovatietaak.

#### *Consequenties voor de studenten*

Voor de psychologiestudenten van de Open Universiteit Nederland betekent het nieuwe statistiekonderwijs dat door de invoering van een vaste langlopende leer- en toepassingslijn de keuze om zelf de volgorde in het bestuderen van modules te bepalen vervalft. Uitstelgedrag in het volgen van statistiekmodules tot vlak voor de masterthesis behoort niet meer tot de mogelijkheden. Dit kan inhouden dat studenten sneller dan voorheen tot de keuze komen de studie al dan niet voort te zetten. Het wordt nu sneller duidelijk wat onderzoek doen in een academische studie inhoudt en dit is

anders dan het stapelen van een aantal losse cursussen. Deze afwijkende benadering binnen het flexibele, modulaire systeem van de Open Universiteit Nederland wordt mogelijk niet door *alle* studenten even enthousiast gewaardeerd.

#### *De onderzoeksopzet*

Het onderzoek naar de vergelijking van het competentiegerichte onderzoeksonderwijs met het traditionele vakgerichte statistiekonderwijs heeft plaatsgevonden aan de hand van een crosssectioneel onderzoek. Een dergelijke opzet is niet optimaal voor het vaststellen van causale relaties en het uitsluiten van de invloeden van 'derde' variabelen. Voor het aantonen van mediatie is de gebruikte onderzoeksopzet ook niet optimaal. Cole en Maxwell (2003) stellen dat men om mediërende processen te kunnen vaststellen men daarvoor minstens 'three waves' aan data nodig heeft zodat relaties vastgelegd kunnen worden tussen de onafhankelijke variabele(n) op tijdstip 1, de mediërende variabele(n) op tijdstip 2 en de uitkomstvariabele(n) op tijdstip 3. De bevindingen die met Structural Equation Modeling in hoofdstuk 5 zijn opgedaan, met name de rol die attitude en motivatie in de twee verschillende leeromgevingen spelen, dienen dan ook als tentatief te worden beschouwd. De longitudinale studie die in hoofdstuk 6 is besproken, kent twee meetmomenten, maar in die studie is maar een deel van het determinantenmodel aan de orde gekomen: motivatie en leerstrategieën. Vanwege de beperkingen in het instrumentarium op tijdstip 1 zijn attitudes buiten beschouwing gebleven en is er in de longitudinale vergelijking noodgedwongen gebruik gemaakt van een matiger kwaliteit aan meetinstrumenten dan in de cross-sectionele vergelijking. Een bijkomende beperking van het longitudinale onderzoek is dat de steekproefgrootte van de groep uit het competentiegerichte onderwijs zeer bescheiden is geweest waardoor het niet verantwoord was in SEM de mediatieprocessen te detecteren. De longitudinale studie moet dan ook worden beschouwd als een exploratief onderzoek gericht op het genereren van onderzoeksvragen voor vervolgonderzoek.

#### *Nonrespons*

Ondanks de lage respons op de uitgezette vragenlijsten en follow-upacties om de respons te bevorderen kan op basis van aanvullend onderzoek worden geconcludeerd dat ten aanzien van het doel van de interventie geen aanwijzingen zijn gevonden die wijzen op vertekening ('nonresponsbias') van de responsgroep (bijlage III.1). Het aanvullende onderzoek toont aan dat er vooral situationele redenen zijn voor niet-deelname (geen tijd, nog niet aan cursus begonnen of gestopt met studie, te lange vragenlijst). Er zijn géén aanwijzingen gevonden dat de lage nonrespons tot vertekening ('nonresponsbias') leidt.

#### *Meetinstrumenten*

De gebruikte meetinstrumenten zijn niet meer volgens de Likertschaalprocedures maar volgens de strengere Raschmodelprocedures vormgegeven. Hierdoor zijn de unidimensionaliteit, betrouwbaarheid en constructvaliditeit van de meetinstrumenten beter gewaarborgd. Echter, dankzij diezelfde Raschmodelprocedures is gebleken dat

meerdere van de meetinstrumenten nog niet optimaal zijn. Vooral de moeilijkheidsgraad van sommige schalen en de spreiding van de items over het onderliggend continuüm laat soms te wensen over.

Een punt dat aandacht verdient, is het instrument dat gebruikt is voor de meting van kennis en begrip van statistiek. De SOLO-taxonomie die gehanteerd is voor de analyse van de toets blijkt zeer goed in staat studenten op een betrouwbare en valide manier in hun niveau van statistisch redeneren te onderscheiden. In samenhang met het Raschmodel zijn studenten en items ook goed te positioneren op het onderliggende continuüm van de schaal voor functionele kennis van de statistiek. Maar het Raschmodel heeft ook laten zien dat niet alle oorspronkelijke items voldoen aan de strenge eisen van eendimensionaliteit, betrouwbaarheid en validiteit. In het longitudinale onderzoek voldoet slechts één item uit de test van  $T_1$  aan de Raschcriteria. Dat item betreft echter de onderlinge relatie tussen drie statistische kernbegrippen. Die drie kernbegrippen staan momenteel in het brandpunt van het onderzoek naar het statistiekonderwijs (Garfield & Ben-Zvi, 2007: 'distribution, centre and variability'). Dankzij het Raschmodel zijn in het longitudinale onderzoek studenten weliswaar goed te positioneren op de schaal voor functionele kennis van de statistiek, maar het is wenselijk de betrouwbaarheid en de validiteit van de verschillende gehanteerde schalen te verhogen door gebruik te maken van meer items.

#### *De onderzoeker als ontwerper, implementeerder en evaluator*

Van den Akker (1998) wijst op een potentieel conflict in rollen als de ontwerper, de implementator en de evaluator dezelfde persoon zijn. Met betrekking tot het ontwerp zijn de denkbeelden voorgelegd en intensief doorgesproken met collega's van de afdelingen Psychologie en het Onderwijstechnologisch Expertisecentrum van de Open Universiteit Nederland. Dit overleg heeft geleid tot het IMTO-project (Van Buuren & Giesbertz, 1998) waarin het ontwerp in een aantal pilots verder is ontwikkeld en beproefd. De meest uitgebreide pilots worden ook in dit proefschrift gepresenteerd (zie hoofdstuk 4). De evaluatie daarvan was in handen van medewerkers van het Onderwijstechnologisch Expertisecentrum. De transformatie van ontwerp naar de diverse onderzoekspractica is grotendeels uitgevoerd door leden van de psychologiestaf die ook het beheer over deze practica hebben gekregen. Bij de evaluatieonderzoeken zijn diverse medeonderzoekers en assistent-onderzoekers betrokken. Door steeds transparante openbaarheid te betrachten – zie Intervention Mapping hierna – teamgericht te werken en in cruciale fasen externe deskundigen te betrekken, is het risico van vermenging van rollen zo minimaal mogelijk gehouden.

### 7.5 **Aanbevelingen voor vervolgonderzoek**

#### *Vervolgonderzoek met een groter bestand aan participanten*

Een groter bestand aan participanten is noodzakelijk om het uitgebreide model, zoals dat is gebruikt in hoofdstuk 5 en 6 van dit proefschrift, met multivariate analysetechnieken te toetsen. De vraag is hoe te komen tot grotere respondentgroepen. Mogelijk

is de omvang van de in dit proefschrift gehanteerde vragenlijst van invloed geweest op de respons. Gebleken is dat bij sommige variabelen een aantal items nagenoeg dezelfde 'moeilijkheidspositie' op het onderliggend continuüm innamen. Hier zou met een minder uitgebreide vragenlijst dezelfde informatie over de variabelen kunnen worden verkregen. Het Raschmodel of de Optimal Design Theory (Berger, 1995, 1996) bieden mogelijkheden om met minder en/of wisselende items te volstaan en toch een beeld te schetsen van de processen die zich op de langere termijn afspelen.

*Vervolgonderzoek van de gehele leer- en toepassingslijn met meer meetmomenten in de bachelorfase*

Vervolgonderzoek met meer meetmomenten en een minder lijvige vragenlijst zou moeten uitwijzen of de positieve trend in de procesvariabelen en de leerresultaten wordt voortgezet en ook waargemaakt wordt in de fase van het schrijven van de bachelorthesis.

*Vervolgonderzoek naar de effectiviteit van competentiegericht onderwijs in andere domeinen*

De resultaten uit de twee evaluatieonderzoeken naar de effecten van het competentiegericht onderwijsontwerp vragen om soortgelijk onderzoek in andere disciplines in het hoger onderwijs die gebruik maken van een competentiegericht onderwijsontwerp. Dergelijk onderzoek dient na te gaan welke onderwijskundige en leertheoretische principes positieve effecten genereren in competentiegericht hoger onderwijs.

*Vervolgonderzoek naar de effectiviteit van de onderzoekscompetentie bij psychologieopleidingen*

Het determinantenmodel beslaat een breed terrein aan thema's dat uitspraken doet over onderwijsleerprocessen en hun opbrengsten. Dat zijn proceskenmerken als attitude, motivatie en zelfgereguleerd leren en leeromgevingkenmerken waarin synergie wordt nagestreefd van inhoud, didactiek en technologie (ICT). Deze thema's staan momenteel in het brandpunt van de onderwijskundige belangstelling, zie bijvoorbeeld het Programma Onderwijsonderzoek 2008-2011 van de NOWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek). Een vervolgonderzoek kan een ondersteuning van de generaliseerbaarheid van het determinantenmodel betekenen. Onderzocht moet worden welke condities (procesvariabelen, leeromgevingkenmerken) de invoering van de onderzoekscompetentie optimaliseren.

*Systematisering en optimalisering van de meetinstrumenten*

Uit de Raschanalyses blijkt dat de dekking van het onderliggende continuüm in enkele gevallen te wensen over laat. Ook het assessment van kennis en begrip van statistiek moet worden verbeterd. Ontwikkelingen in bijvoorbeeld Optimal Design Theory maken het mogelijk meetinstrumenten kritischer te inspecteren op hun meetkwaliteiten en optimale instrumenten te ontwerpen (zie onder anderen Berger, 1996).



*Samenwerkend leren*

In de pilots van het IMTO-project, zijn eerste ervaringen opgedaan met samenwerkend leren in een elektronische leeromgeving. Bij de implementatie van het ontwerp in het curriculum is echter niet gekozen voor de inbedding van de onderzoekscompetentie in een leeromgeving waarin samenwerking mogelijk is. Redenen waren de toenmalig ervaren ontoereikendheid van de ICT-voorzieningen en het grote beroep dat werken in een elektronische leeromgeving op de begeleidings- en studietijd legt. Gezien de stand van zaken met ICT op dit moment, de ervaringen die in tal van onderzoeken met e-learning zijn opgedaan en gegeven het feit dat de implementatie van het competentiegericht onderwijs in het curriculum nagenoeg is voltooid, moeten experimenten met samenwerkend leren weer op de onderzoeksagenda komen te staan. Een uitdaging is hierbij de verscheidenheid aan groepen die onder de studenten van de Open Universiteit Nederland zijn te onderscheiden. Niet alle studenten die hebben gekozen voor de Open Universiteit Nederland zullen kunnen of willen samenwerken. Er zijn studenten die het aangenaam vinden om met elkaar te kunnen samenwerken. Voor afstandsstudenten kan een elektronische leeromgeving uitkomst bieden. Een andere groep zou passief willen deelnemen, door berichten te lezen in de nieuwsgroepen en zich zodoende een beeld te vormen (vergelijk sites waarin consumenten ervaringen kunnen uitwisselen). Weer een andere groep zal wel willen samenwerken, maar heeft daar verder geen mening over. En een laatste groep kiest voor de Open Universiteit Nederland omdat ze dan geheel in eigen tempo kan studeren.

## 7.6 **Aanbevelingen voor de praktijk van statistiekonderwijs nieuwe stijl**

Mocht vakoverstijgende integratie van statistiek en methoden en technieken van onderzoek, zoals bij de opleiding Psychologie van de Open Universiteit Nederland organisatorisch lastig zijn, dan kan geadviseerd worden de cursussen statistiek meer in te bedden in het domein waarin de cursussen worden gegeven en de hoeveelheid leerstof te bepalen tot enkele veel toegepaste kernbegrippen. Beter enkele veel toegepaste begrippen goed uitgewerkt die én geleerd én toegepast worden op hun studieterrein dan cognitieve overbelasting door een waaier van begrippen die relatief betekenisloos zijn voor de doelgroep. Dit voorkomt dat studenten blijven steken in de eerste declaratieve opsommingsfase die Anderson (2000) onderscheidt in het proces van leek tot expert. Een klassiek alternatief is het spreiden van de leerstof in een aantal korte cursussen met veel herhalingen over een langere periode zodat bij de bachelorthesis niet alles is weggezaakt.

## 7.7 **Tot slot**

De evaluatie heeft zich beperkt tot het eerste practicum. Er zullen nog zes geïntegreerde onderzoekspractica volgen waarin statistiek steeds weer opnieuw wordt toegepast. De opleiding Psychologie van de Open Universiteit Nederland heeft daarmee een eerste stap gezet om de onderzoekscompetentie vorm te geven. Integratie van statistiek, M&T



en psychologie is tegelijk een uitdagende stap. Integratie opent het perspectief voor nauwe samenwerking en afstemming tussen docenten van verschillende disciplines en voor meer begrip voor elkaars vak. Het vak statistiek heeft nu duidelijker dan voorheen een meer verborgen (want geïntegreerde) én een meer prominente (want overal aanwezige) plaats in het curriculum gekregen.

*Bijlage I.0*

## **Stappen Intervention Mapping**

Cel	A	B	C	D																								
1	Hoofdstuk	1 Probleem → Interventiedoel	2 Interventiedoel → Veranderdoelen	3 Veranderdoelen → Instructieaanwijzingen																								
2	Onderzoeksvraag	Wat zijn de problemen in het statistiekonderwijs en welke verklaringen worden daarvoor gegeven?	Welke determinanten bepalen het succes van het leren van statistiek?	Welke onderwijskundige instructieaanwijzingen zijn effectief om de gewenste veranderingen in de leeromgeving te bewerkstelligen die leiden tot succes van het leren van statistiek?																								
3	Intervention Mapping taken	Taak 1.1 Analyseer de problemen en mogelijkheden voor oplossingen. Taak 1.2 Formuleer gewenste programmauitkomsten ('Interventiedoel').	Taak 2.1 Specificeer leeruitkomsten in gedragsdoelen Taak 2.2 Selecteer belangrijke en veranderbare determinanten van die gedragingen. Taak 2.3 Maak een matrix van veranderdoelen door gedragsdoelen te kruisen met determinanten van de gedragingen.	Taak 3.1 Zoek theorieën bij de veranderdoelen. Taak 3.2 Leid methodieken af uit de theorie en de onderzoeksliteratuur. Taak 3.3 Vertaal de methodieken in praktische technieken. Taak 3.4 Ga na of de technieken overeenstemmen met de veranderdoelen																								
4	Theorie	<table><tr><th>Problemen Statistiekonderwijs</th><th>Doel van interventie</th></tr><tr><td>- Studentkenmerken</td><td>- indirecte aanpak</td></tr><tr><td>- Onderwijskenmerken</td><td>- indirecte aanpak</td></tr><tr><td>- Leeromgevingskenmerken</td><td>- directe aanpak</td></tr></table>	Problemen Statistiekonderwijs	Doel van interventie	- Studentkenmerken	- indirecte aanpak	- Onderwijskenmerken	- indirecte aanpak	- Leeromgevingskenmerken	- directe aanpak	Matrix van veranderdoelen (bijlage I.1) <table><tr><th colspan="2">Determinanten</th></tr><tr><th>Externe</th><th>Persoonlijke</th></tr><tr><td>Inhoud</td><td>Attitude</td></tr><tr><td>Didactiek</td><td>Motivatie</td></tr><tr><td>Technol.</td><td>Leerstrategieën</td></tr><tr><th>Gedragsdoelen</th><th>Veranderdoelen</th></tr></table>	Determinanten		Externe	Persoonlijke	Inhoud	Attitude	Didactiek	Motivatie	Technol.	Leerstrategieën	Gedragsdoelen	Veranderdoelen	bijlage I.2 <table><tr><th></th><th>Methodieken &amp; praktijken</th></tr><tr><th>Veranderdoelen</th><th>Instructieaanwijzingen</th></tr></table>		Methodieken & praktijken	Veranderdoelen	Instructieaanwijzingen
Problemen Statistiekonderwijs	Doel van interventie																											
- Studentkenmerken	- indirecte aanpak																											
- Onderwijskenmerken	- indirecte aanpak																											
- Leeromgevingskenmerken	- directe aanpak																											
Determinanten																												
Externe	Persoonlijke																											
Inhoud	Attitude																											
Didactiek	Motivatie																											
Technol.	Leerstrategieën																											
Gedragsdoelen	Veranderdoelen																											
	Methodieken & praktijken																											
Veranderdoelen	Instructieaanwijzingen																											
5	Empirisch	<p>Determinantenmodel (fig. 2.1)</p>																										

E	F	G																															
4 Instructieaanwijzing → Ontwerp	5/6 Ontwerp → Implementatie	7 Implementatie → Evaluatie																															
<i>Hoe kunnen de instructieaanwijzingen worden gesynthetiseerd tot een competentiegericht ontwerp dat een effectieve leeromgeving voor de onderzoekscompetentie inhoudt en hoe kan dat competentiegericht ontwerp worden geïmplementeerd in een bestaand, vakgericht curriculum?</i>	<i>In welke mate is de competentiegerichte leeromgeving effectiever in het realiseren van leerprocessen en leeruitkomsten dan de vakgerichte leeromgeving, ook op langere termijn?</i>	<i>In hoeverre leidt competentiegericht onderwijs (de interventie) tot een effectieve leeromgeving waarin de afstand tussen gewenste en gerealiseerde leerdoelen ten aanzien van begrip en toepassing van de statistiek in psychologisch onderzoek kleiner is dan in het vakgerichte onderwijs?</i>																															
<p>Taak 4.1 Combineer de verschillende technieken in één plan</p> <p>Taak 4.2 Ontwikkel prototypen voor onderwijsleeromgeving: scope, sequentie van taken, thema's (→ competentiekaart)</p> <p>Taak 4.3 Test het ontwerp in pilots.</p>	<p>Taak 5.1 Bepaal de implementatiestrategie.</p> <p>Taak 5.2 Test de materialen uit bij de doelgroep en vergelijk de resultaten met de oude situatie.</p>	<p>Taak 6.1 Beschrijf het programma.</p> <p>Taak 6.2 Beschrijf onderzoeksvragen voor programma-uitkomsten en proces.</p>																															
<p>4C/ID: Tien ontwikkelstappen (bijlage I.3)</p> <p>Onderwijsblauwdruk onderzoekscompetentie (bijlage I.4)</p> <table><tr><th>Comp. niveau</th><th>Fase 1 Inleiding</th><th>Fase 2 Methoden</th><th>Fase 3 Resultaten</th><th>Fase 4 Discuss</th></tr><tr><td>1</td><td></td><td></td><td></td><td></td></tr><tr><td>2</td><td></td><td></td><td></td><td></td></tr><tr><td>3</td><td></td><td></td><td></td><td></td></tr><tr><td>etc.</td><td></td><td></td><td></td><td></td></tr></table>	Comp. niveau	Fase 1 Inleiding	Fase 2 Methoden	Fase 3 Resultaten	Fase 4 Discuss	1					2					3					etc.					<p>Tabel 5.2 (p. 81)</p> <table><tr><th colspan="2">Statistiekonderwijs</th></tr><tr><th>Vakgericht (VAK)</th><th>Competentiegericht (COMP)</th></tr><tr><td><p>Kenmerken</p><ul style="list-style-type: none"><li>- Doelstelling</li><li>- Instructie</li><li>- Inhoud</li><li>- ECTS-eenheden</li><li>- Leermaterialen</li><li>- Toetsing</li></ul></td><td><p>Kenmerken</p><ul style="list-style-type: none"><li>- Doelstelling</li><li>- Instructie</li><li>- Inhoud</li><li>- ECTS-eenheden</li><li>- Leermaterialen</li><li>- Toetsing</li></ul></td></tr></table>	Statistiekonderwijs		Vakgericht (VAK)	Competentiegericht (COMP)	<p>Kenmerken</p> <ul style="list-style-type: none"><li>- Doelstelling</li><li>- Instructie</li><li>- Inhoud</li><li>- ECTS-eenheden</li><li>- Leermaterialen</li><li>- Toetsing</li></ul>	<p>Kenmerken</p> <ul style="list-style-type: none"><li>- Doelstelling</li><li>- Instructie</li><li>- Inhoud</li><li>- ECTS-eenheden</li><li>- Leermaterialen</li><li>- Toetsing</li></ul>	<p>1 Optimaliseren ontwerp</p> <p>2 Model</p> <p>3 Reikwijdte onderzoek</p> <p>4 Nader onderzoek</p> <p>5 Samenwerkend leren</p> <p>6 Intervention Mapping</p> <p>7 Onderzoeker als ontwerper/evaluator</p>
Comp. niveau	Fase 1 Inleiding	Fase 2 Methoden	Fase 3 Resultaten	Fase 4 Discuss																													
1																																	
2																																	
3																																	
etc.																																	
Statistiekonderwijs																																	
Vakgericht (VAK)	Competentiegericht (COMP)																																
<p>Kenmerken</p> <ul style="list-style-type: none"><li>- Doelstelling</li><li>- Instructie</li><li>- Inhoud</li><li>- ECTS-eenheden</li><li>- Leermaterialen</li><li>- Toetsing</li></ul>	<p>Kenmerken</p> <ul style="list-style-type: none"><li>- Doelstelling</li><li>- Instructie</li><li>- Inhoud</li><li>- ECTS-eenheden</li><li>- Leermaterialen</li><li>- Toetsing</li></ul>																																
<p>Eerste kwalitatieve bevindingen werkbaarheid ontwerp in IMTO-pilots (tabellen 4.4 - 4.10) (p. 67 e.v.)</p> <table><tr><th>Pilots</th><th>Werkbaarheid ontwerp</th></tr><tr><td>IMTO-bètarun IMTO-5</td><td>bevindingen</td></tr></table>	Pilots	Werkbaarheid ontwerp	IMTO-bètarun IMTO-5	bevindingen	<p>Rasch-schalen</p> <p>MANOVA/ANCOVA</p> <p>Structural Equation Modeling</p> <p>Standardized estimates chi square=172,416 df=82 p&lt;.000</p>	<p>1 Optimaliseren meetinstrumenten</p> <p>2 Onderzoeksopzet</p> <p>3 Longitudinale studies (vervolgonderzoeken)</p> <p>4 Nonrespons</p> <p>5 Generaliseerbaarheid</p>																											
Pilots	Werkbaarheid ontwerp																																
IMTO-bètarun IMTO-5	bevindingen																																



## Referenties

- Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Allport, G.W. (1935). Attitudes. In C. Murchinson (Ed.), *A Handbook of Social Psychology* (pp. 798-844). Worcester: Clark University Press.
- American Psychological Association (2001). *Publication Manual of the American Psychological Association*, 5<sup>th</sup> ed. Washington: APA.
- Anderson, J.R. (2000). *Cognitive Psychology and its Implications* (5th ed.). New York: Worth Publishers.
- Arbuckle, J.L. (2003). *Update to the AMOS User's Guide*. Chicago: SmallWaters Corporation.
- Attema, N. (1996). *Evaluatieverslag propeadeuse statistiek onderwijs 1995-1996*. Leiden: Universiteit van Leiden.
- Ausubel, D.P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 51, 267-272.
- Ausubel, D.P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart & Winston.
- Babbie, E. (2002). *The Practice of Social Research* (6th ed.). Belmont, CA: Wadsworth.
- Bakker, A. (2004). *Design research in statistics education: on symbolizing and computer tools*. University of Utrecht, Utrecht.
- Baloğlu, M. (2001). *An application of structural equation modeling techniques in the prediction of statistics anxiety among college students*. Unpublished doctorate dissertation, Texas A & M University.
- Baloğlu, M. (2003). Individual differences in statistics anxiety among college students. *Personality and Individual Differences*, 34, 855-865.
- Baloğlu, M., & Zelhart, P. F. (2003). Statistical Anxiety: A Detailed Review of the Literature. *Psychology and Education*, 40(2), 27-37.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1993). Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*, 28(2), 117-148.
- Bandura, A. (1997). Self-efficacy. *Harvard Mental Health Letter*, 13(9).
- Baron, R.M., & Kenny, D.A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.
- Bartholomew, L.K., Parcel, G.S., Kok, G., & Gotlieb, N. H. (2006). *Planning health promotion programs: an Intervention Mapping approach*. San Francisco: Jossey-Bass.
- Beenen, F., & Van der Werff, J.J. (1967). Statistiefobie. Een onderzoek naar de weerstanden tegen statistiek. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, 22(4), 209-233.
- Ben-Zvi, D., & Garfield, J. (Eds.). (2004). *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Deventer: Kluwer Academic Publishers.
- Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, 107(2), 238-246.

- Berger, M.P.F. (1996). Evolutie van de MenS. *Ontwikkelingen van het vakgebied Methodologie en Statistiek*. Maastricht: Universitaire Pers Maastricht.
- Berger, M.P.F., & Van der Linden, W. (1995). Het ontwerpen van tests met verschillende optimaliteitscriteria. *Tijdschrift voor Onderwijsresearch*, 20(1), 79-92.
- Berk, R.A. (1996). Student Ratings for 10 Strategies for Using Humor in College Teaching. *Journal on Excellence in College Teaching*, 7(3), 71-92.
- Biggs, J.B. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364.
- Biggs, J.B. (2003). *Teaching for Quality Learning at University*. Buckingham: SRHE and Open University Press.
- Biggs, J.B., & Collis, K.F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. New York, NY: Academic press.
- Biggs, J.B., & Collis, K.F. (1989) Towards a model of school-based curriculum development and assessment using the SOLO taxonomy, *Australian Journal of Education*, 33, 151-163.
- Bijker, M.M., & Van Buuren, J.A. (2006). *Effectmeting van competentiegericht onderwijs bij psychologiestudenten in afstandsonderwijs. Effecten op motivaties, attitudes, leerstrategieën en kennis*. Paper presented at the 33ste Onderwijs Research Dagen, Amsterdam, Nederland.
- Bijker, M.M., Wynants, G., & Van Buuren, J.A. (2006). *A comparative study of the effects of motivational and attitudinal factors on studying statistics*. Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Bahia, Brazilië.
- Bland, J. M. & Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, 310, 170.
- Blom, S., Severiens, S., Broekkamp, H., & Hoek, D. (2005). *Zelfstandig leren van allochtone en autochtone kinderen in het Studiehuis*. Amsterdam: Instituut voor de Lerarenopleiding.
- Blom, S.V., Hoek, D.J., & Ten Dam, G.T.M. (2007). Metacognitieve zelfregulatie, motivatie en perceptie van klassenklimate. Zijn er sociale verschillen? *Pedagogische Studiën*, 84, 20-36.
- Bond, T.G., & Fox, C.M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New Jersey: Lawrence Erlbaum Associates.
- Boulton-Lewis, G. (1994) Tertiary students' knowledge of their own learning and a SOLO taxonomy, *Higher Education*, 28, 387-402.
- Boulton-Lewis, G. (1995) The SOLO taxonomy as a means of shaping and assessing learning in higher education, *Higher Education Research and Development*, 14, 143-154.
- Boulton-Lewis, G. (1998) Applying the SOLO taxonomy to learning in higher education, in: B. Dart & G. Boulton-Lewis (Eds) *Teaching and Learning in Higher Education* (pp. 201-221). Melbourne: ACER.
- Boulton-Lewis, G., & Dart, B. (1994) Assessing students' knowledge of learning: a comparison of data collection methods. In G. Gibbs (Ed.), *Improving Student Learning: Theory and Practice* (pp. 263-277). Oxford: OCSA.
- Boulton-Lewis, G., Pillay, H., & Wilss, L. (2004). *Teaching for diversity in higher education in Australia: proposed research*. Brisbane: School of learning and Professional Studies, Faculty of Education, QUT.
- Bradstreet, T.E. (1996). Teaching Introductory Statistics Courses So That Nonstatisticians Experience Statistical Reasoning. *The American Statistician*, 50, 69-78.
- Bransford, J.D., Brown, A.L., & Cocking, R.R. (2000). *How People Learn: Brain, Mind, Experience, and School Committee on Developments in the Science of Learning*. Washington: National Academy Press.
- Brehm, J. (1993). *The phantom respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.
- Britz, G., Emerling, D., Hare, L., Hoerl, R. W., & Shade, J. (1997). How to Teach Others to Apply Statistical Thinking. *Quality Progress*(June), 67-79.



- Brophy, J. (1999). *Teaching* (Vol. 1). International Academy of Education.
- Bross, I.D.J. (1974). The role of the statistician: scientist or shoe clerk. *The American*
- Brown, J.S., Collins, A., & Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
- Brown, K.W., & Ryan, R.M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84(4), 822-848.
- Burnett, P.C. (1999). Assessing the structure of learning outcomes from counseling using the SOLO taxonomy: an exploratory study. *British Journal of Guidance & Counseling*, 27(4), 567-581.
- Burton, R.F. (2004). Can item response theory help us improve our tests? *Medical Education*, 38, 336-339.
- Butler, D.L., & Winne, P.H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Butler, R.S. (1998). On the failure of the widespread use of statistics. *Amstat News*, 251(84).
- Byrne, B.M. (2001). *Structural Equation Modeling with Amos. Basic concepts, applications and programming*. New Jersey: Lawrence Erlbaum Associates.
- Callingham, R. & Watson, J.M. (2005). Measuring Statistical Literacy. *Journal of Applied Measurement*, 6(1), 19-47.
- Campbell, K.J., Watson, J.M., & Collis, K.F. (1992). Volume Measurement and Intellectual Development. *Journal of Structural Learning*, 11(3), 279-298.
- Chalmers, A.D., & Hill, B. (1973). The Numeracy of Social Science Students. *International Journal of Mathematical Education in Science and Technology*, 4, 17-24.
- Chan, C.C., Tsui, M.S., Chan, Y.C., & Hong, J.H. (2002). Applying the Structure of the Observed Learning Outcomes (SOLO) Taxonomy on Student's Learning Outcomes: an empirical study. *Assessment & Evaluation in Higher Education*, 27(6).
- Chance, B.L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education*, 10(3).
- Chen, C.S. (2002). Self-regulated learning strategies and achievement in an introduction to information systems course. *Information Technology, Learning, and Performance Journal*, 20(1), 11-25.
- Clark, J., Karuat, G., Mathews, D., & Wimbish, J. (2003). *The Fundamental Theorem of Statistics: Classifying Student understanding of basic statistical concepts*. [On-line]
- Cobb, G.W. (1992). *Teaching statistics* (No. 22): MAA Notes No. 22.
- Cobb, G.W. (1993). Reconsidering statistics education: A national science foundation conference. *Journal of Statistics Education*, 1(1), on-line.
- Cohen, J.W. (1988). *Statistical power analysis for the behavioural sciences*. (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cole, D.A., & Maxwell, S.E. (2003). Testing Mediation Models With Longitudinal Data: Questions and Tips in the Use of Structural Equation Modeling. *Journal of Abnormal Psychology*, 112(4), 558-577.
- Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing and mathematics. In L.B. Resnick (ed.). *Knowing, learning and instruction. Essays in honor of Robert Glaser* (pp. 453-494). Hillsday, N.J.: Lawrence Erlbaum.
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 6(11), 38-46.
- Commissie Levelt (2006). *Reguliere masteropleidingen psychologie in Nederland: is een tweejarig curriculum noodzakelijk?* Leuven: Nederlands-Vlaamse Accreditatie Organisatie NVAO.
- Covington, M.V. (2000). Goal theory, motivation, and school achievement: an integrative review. *Annual Review of Psychology*, 51, 171-200.

- Curda, L. K. (1997). *Validation of a measure of teachers' efficacy and outcome expectations in the content domains of reading and mathematics*. Norman, Oklahoma: University of Oklahoma.
- Curda, S.K. (1997). *Learners' motivational characteristics in statistics: a causal model*. Norman, Oklahoma: University of Oklahoma.
- Czikszentmihalyi, M. (1990). *Flow: the classic work on how to achieve happiness*. London: Random House Group Ltd.
- Dahlgren, L.-O. (1997) Learning conceptions and outcomes, in F. Marton, Dhounsell & N. Entwistle (Eds.) *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (2<sup>nd</sup> ed., pp. 23-38). Edinburgh: Scottish Academic Press.
- Darabi, A.A. (2005). Application of Cognitive Apprenticeship Model to a Graduate Course in Performance Systems Analysis: A Case Study. *Educational Technology Research & development*, 53(1), 49-61.
- Dauphinee, T.L., Schau, C., & Stevens, J.J. (1997). Survey of Attitudes Toward Statistics: Factor Structure and Factorial Invariance for Women and Men. *Structural Equation Modeling*, 4(2), 129-141.
- De Groot, A.D. (1961). *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen*. Den Haag: Mouton.
- delMas, R., Garfield, J., & Chance, B.L. (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3).
- Derry, S.J., Levin, J.R., Osana, H., Jones, M.S, & Peterson, M. (2000). Fostering students' statistical and scientific thinking: Lessons learned from an innovative college course. *American Educational Research Journal*, 37(3), 747-773.
- Dillon, K.M. (1982). Statisticophobia. *Teaching of Psychology*, 9, 117.
- Dolmans, D.H.J.M., & Schmidt, H.G. (1996). Inzichten over leren en implicaties voor het onderwijs. *Bulletin Medisch Onderwijs*, 15(3), 114-119.
- Donald, J.G. (2002). *Learning to think*. San Francisco: Jossey-Bass.
- Downing, S.M. (2003). Item response theory: application of modern test theory in medical education. *Medical Education*, 37, 739-745.
- Eagly, A.H., & Chaiken, S. (1993). *The Psychology of Attitudes*. Orlando: Harbourt Brace College Publishers.
- Eccles, J.S. (1983). Expectancies, Values and Academic Behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motives* (pp. 75-146). San Francisco: W.H. Freeman and Company.
- Eccles, J.S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. In S.T. Fiske, D. L. Schacter & C. Sahn-Waxler (Eds.), *Annual Review of Psychology* (109-132). Palo Alto, CA.
- Ellemers, J.E. (2004). Wat we nog niet weten. Tekortkomingen en blinde vlekken bij de analyse van de Nederlandse maatschappij. *Sociologische Gids*, 51(4), 482-487.
- Elshout-Mohr, M., & Oostdam, R. (2001). *Assessment van competenties in een dynamisch curriculum*. Amsterdam: Pierson Révész Bibliotheek.
- Engels-Freeke, M., & Flach, D. P. (1999). Dr. Stat, a computerized statistical aid for psychology students. *Kwantitatieve Methoden*, 20(60), 41-54.
- Entwistle, N. J. (2000). *Constructive alignment improves the quality of learning in higher education*. Paper gepresenteerd op de Onderwijs Research Dagen, Leiden.
- Felder, R.M., & Brent, R. (2001). Effective strategies for cooperative learning. *Journal of Cooperation & Collaboration in College teaching*. 10(2), 69-75.
- Gal, I., & Garfield, J. (Eds.). (1997). *The Assessment Challenge in Statistics Education*. Amsterdam: IOS Press and the International Statistical Institute.

- Gal, I., & Ginsburg, L. (1994). The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework. *Journal of Statistics Education* 2(2).
- Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J.B. Garfield (Eds.) *The assessment challenge in statistics education* (pp. 37-51). Netherlands: IOS Press.
- Garfield, J. B. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, 1(1).
- Garfield, J. (1997). Discussion. *International Statistical Review*, 65(2), 137-141.
- Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, 10(3).
- Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, Vol 19(1), 44-63.
- Garfield, J., & Ben-Zvi, D. (2007). How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics. *International Statistical Review*, 75(3), 372-396.
- Garfield, J., delMas, R., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. Lovett & P. Shah (Eds.), *Thinking with Data* (pp. 117-148). Mahwah, NJ: Lawrence Erlbaum.
- Gelso, C. J. (1979). Research in counseling: Methodological and professional issues. *The Counseling Psychologist*, 5(3), 7-35.
- Gelso, C. J. (2006). On the Making of a Scientist-Practitioner: A Theory of Research Training in Professional Psychology. *Training and Education in Professional Psychology*, 5(1), 3-16.
- Giesbertz, W., & Van Buuren, H. (2004). Competentiebegrip bij de OUNL. Van hype naar realistische praktijk. In L. Dercksen, H. Dekker, M. Van der Klink & S. Wagenaar (Eds.), *Competentiegericht leren: de context centraal*. (5(3), pp. 80-88). Doetinchem: Reed Business Infomation.
- Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistical Education*, 5(3).
- Goldberg, L.R. (1990). An alternative description of personality: the Big Five Factor Structure. *Journal of Personality and Social Psychology*, 59(10), 1216-1229.
- Gordon, S. (1999). *An instrument for exploring students' approaches to learning statistics*. Paper presented at the the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Groves, R.M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Harlow, L.L., Burkholder, G.J., & Morrow, J.A. (2002). Evaluating attitudes, skill and performance in a learning-enhanced quantitative method course: a structural modelling approach. *Structural Equation Modelling*, 9(3), 413-430.
- Hattie, J., & Purdie, N. (1998). The SOLO model: adressing fundamental measurement issues. In B. C. Dart & G. M. Boulton-Lewis (Eds.), *Teaching and Learning in Higher Education* (pp. 145-176). Melbourne: ACER.
- Hendricks, C.C. (2001). Teaching causal reasoning through cognitive apprenticeship: What are results from situated learning? *The Journal of Educational Research*, 94(5), 302.
- Hilton, S.C., Schau, C.G., & Olsen, J.A. (2004). Survey of Attitudes Toward Statistics: Factor Structure Invariance by Gender and by Administration Time. *Structural Equation Modeling*, 11(1), 92-109.
- Hogg, R.V. (1991). Statistical Education: Improvements Are Badly Needed. *The American Statistician*, 45(4), 342-343.
- Hogg, R.V. (1992). Workshop on Statistical Education. In L. Steen (ed.), *Heeding the Call for Change. Suggestions for Curricular Action*. MAA Notes No. 22, Washington: Mathematical Association of America.

- Holmes, K. (2005) Analysis of Asynchronous Online Discussion using the SOLO Taxonomy. *Australian Journal of Educational & Developmental Psychology*, 5, 117-127.
- Hommel, M.A. (2006). *Zelfinstructie bij gesprekst raining voor afstandsonderwijs. Effecten op vaardigheid, self-efficacy, motivatie en transfer*. Rotterdam: Erasmus Universiteit Rotterdam.
- Hoogveld, A.W.M. (2003). *The Teacher as Designer of Competency-Based Education*. Heerlen: Open Universiteit Nederland.
- Houkes, I. (2002). *Work and individual determinants of intrinsic work motivation, emotional exhaustion and turnover intention*. Universiteit van Maastricht, Maastricht.
- Hutjes, J.M., & Van Buuren, J.A. (1992). *De gevalsstudie. Strategie van kwalitatief onderzoek*. Meppel/Amsterdam: Boom.
- Imbos, T., & Dekeyser, H. (1999). Methodologische en Statistische Educatie als service onderwijs: Problemen en mogelijke oplossingen. *Kwantitatieve Methoden*, 60 (Februari 1999), 9-27.
- Jackson, J.W. (2002). Enhancing self-efficacy and learning performance. *The Journal of Experimental Education*, 70(3), 243-254.
- Janssen-Noordman, A.M.B., & Van Merriënboer, J.J.G. (2002). *Innovatief onderwijs ontwerpen. Via leertaken naar complexe vaardigheden*. Groningen: Wolters-Noordhoff.
- Janssens, S., Verschaffel, L., De Corte, E., Elen, J., Struyf, E., Van Damme, J. & Vandenberghe, R. (2003). *Didactiek in beweging*. Antwerpen: Wolters Plantyn.
- Jochems, W.M.G., Van Merriënboer, J.G., & Koper, R. (2004). *Integrated e-learning. Implications for pedagogy, technology & organization*. London: Routledge Falmer.
- Johnson, R.T., Johnson, D.W., & Stanne, M.B. (2000). *Cooperative learning methods: A meta-analysis*. Minneapolis, MN: University of Minnesota.
- Jones, G.A., Langrall, C.W., Thornton, C.A., & Mogill, A.T. (1997). A framework for assessing and nurturing children's thinking in probability. *Educational Studies in Mathematics*, 32, 101-125.
- Jones, D.R. & Rushton, L. (1982). Simultaneous Inference in Epidemiological Studies. *International Journal of Epidemiology*, 11(3), 276-282.
- Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and Testing Mediation and Moderation in Within-Subject Designs. *Psychological Methods*, 6(2), 115-134.
- Kalton, G. (1973). Problems and Possibilities with Teaching Introductory Statistics to Social Scientists. *International Journal of Mathematical Education in Science and Technology*, 4, 7-16.
- Karabenick, S.A., & Knapp, J.R. (1991). Relationship of Academic Help Seeking to the Use of Learning Strategies and Other Instrumental Achievement Behavior in College Student. *Journal of Educational Psychology*, 83(2), 221-230.
- Keeler, C.M., & Steinhurst, R.K. (1995). Using small groups to promote active learning in the introductory statistics course: a report from the field. *Journal of Statistics Education*, 3(2).
- Kember, D., Wong, A., & Leung, D.Y.P. (1999). Reconsidering the dimensions of approaches to learning. *British Journal of Educational Psychology*, 69, 323-343.
- Kirschner, P., Van Vilteren, P., Hummel, H., & Wigman, M. (1997). The Design of a Study Environment for Acquiring Academic and Professional Competence. *Studies in Higher Education*, 22(2), 151-171.
- Kneefel, B., & Leeden, R. v. d. (1999). Computerondersteund statistiekonderwijs aan eerstejaars psychologiestudenten in Leiden. *Kwantitatieve Methoden*, 60 (Februari 1999), 55-69.
- Konold, C. (1989). Informal Conceptions of Probability. *Cognition and Instruction*, 6, 59-98.
- Kreijns, K. (2004). *Sociable CSCL Environments: Social Affordances, Sociability, and Social Presence*. Heerlen: Open Universiteit Nederland.
- Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.

- Lagerwaard, H., & Mul, J. (1982). *Scripties onderzocht. Een analyse van de kwaliteit van doctoraalscripties in de culturele antropologie, de politicologie en de sociologie en de tijd, die het schrijven ervan kost, met aanbevelingen ter verbetering*. Leiden: Centrum voor Onderzoek van Maatschappelijke Tegenstellingen.
- Lake, D. (1999). Helping students to go SOLO: teaching critical numeracy in the biological sciences. *Journal of Biological Education*, 33(4), 191-198.
- Lalonde, R.N., & Gardner, R.C. (1993). Statistics as a second Language? A Model for Predicting Performances in Psychology Students. *Canadian Journal of Behavioural Science*, 25(1), 108-125.
- Lane, D. M., & Peres, S. C. (2006). *Interactive simulations in the teaching of statistics: Promise and pitfalls*. Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Brasil.
- Lane, D. M., & Tang, Z. (2000). Effectiveness of simulation training on transfer of statistical concepts. *Journal of Educational Computing Research*, 22(4), 383-396.
- Lave, J. (1988). *Cognition in Practice: Mind, mathematics and culture in everyday life*. New York: Cambridge University Press.
- Lave, J., & Wenger, E. (1991). *Situated Learning. Legitimate peripheral participation*, Cambridge: University of Cambridge Press.
- Lewalter, D., & Krapp, A. (2004). The role of contextual conditions of vocational education for motivational orientations and emotional experiences. *European Psychologist*, 9(4), 210-221.
- Linacre, J.M. (2005). *A User's Guide to WINSTEPS Ministep Rasch-model Computer Programs*. Chicago: MESA Press.
- Linnenbrink, E.A., & Pintrich, P.R. (2002a). Achievement goal theory and affect: an asymmetrical bidirectional model. *Educational Psychologist*, 37(2), 69-78.
- Linnenbrink, E.A., & Pintrich, P.R. (2002b). Motivation as an Enabler for Academic Success. *School Psychology Review*, 31(3), 313-327.
- Linnenbrink, E.A., & Pintrich, P.R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*(19), 119-137.
- Little, T.D., Card, N.A., Bovaird, J.A., Preacher, K.J., & Crandall, C.S. (2007). Structural Equation Modeling of Mediation and Moderation With Contextual Factors. In T. D. Little, J. A. Bovaird & N. A. Card (Eds.), *Modeling Contextual Effects in Longitudinal Studies* (Vol. 9, pp. 207-230): Psychology Press.
- Lonka, K., Olkinuora, E., & Mäkinen, J. (2004). Aspects and prospects of Measuring Studying and Learning in Higher Education. *Educational Psychology Review*, 16(4), 301-323.
- Lovett, M. (2001). A collaborative convergence on studying reasoning processes: A case study in statistics. In D. Klahr & S. Carver (Eds.), *Cognition and Instruction: Twenty-Five Years of Progress* (pp. 347-384). Mahwah, NJ: Lawrence Erlbaum.
- Lovett, M.C., & Greenhouse, J.B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician*, 54(3), 196-206.
- MacKinnon, D.P., Fairchild, A.J., & Fritz, M.S. (2007). Mediation Analysis. *Annual Review of Psychology*, 58, 593-614.
- Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistical Education*, 6(3).
- Martens, R.L., Gulikers, J., & Bastiaens, T.J. (2004). The impact of intrinsic motivation on e-learning in authentic computer tasks. *Journal of Computer Assisted Learning*, 20, 368-376.
- Mathews, D., & Clark, J. (2003). *Successful Students' Conceptions of Mean, Standard Deviation and the Central Limit Theorem*.
- McKeachie, W.J., & Hofer, B.K. (2001). *Teaching tips: Strategies, research, and theories for college and university teachers* (11th ed.). Houghton: Mifflin Company

- Mellenbergh, G.J. (1980). Theorie op verschillende niveaus. *Nederlands Tijdschrift voor de Psychologie*, 35, 275-288.
- Merrill, M.D. (2002). First Principles of Instruction. *Educational Technology Research and Development*, 50(3), 43-59.
- Middleton, J.A., & Spanias, P.A. (1999). Motivation for Achievement in Mathematics: Findings, Generalizations, and Criticisms of the Research. *Journal for Research in Mathematics Education*, 30(1), 65-88.
- Midgley, C., Kaplan, A., Middleton, M., & Maehr, M.L. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology*, 23, 113-131.
- Miles, M.B., & Huberman, A.M. (1984). *Qualitative Data Analysis. A Sourcebook of New Methods*. Beverly Hills: Sage publications.
- Miles, M.B., & Huberman, A.M. (1994). *Qualitative Data Analysis. An Expanded Sourcebook* (2 ed.). Thousand Oaks: Sage Publications.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Moonen, J.C. (1978). *Computergestuurd onderwijs: een onderzoek naar de mogelijkheden tot geïntegreerd gebruik van een computergestuurd systeem in een statistiekkurriculum*. Leiden: Rijksuniversiteit Leiden.
- Moore, D.S. (1997). New pedagogy and new content: the case of statistics. *Statistical review*, 65(2), 123-165.
- Morris, J. (2001). The conceptions of the nature of learning of first-year physiotherapy students and their relationship to students' learning outcomes. *Medical Teacher*, 23(5), 503-507.
- Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education*, 12(1).
- Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition*, 19(6), 543-557.
- Newman, R.S. (1990). Children's Help-Seeking in the Classroom: The Role of Motivational Factors and Attitudes. *Journal of Educational Psychology*, 82, 71-80.
- Nicholls, D. F. (2001). Future Directions for the Teaching and Learning of Statistics at the Tertiary Level. *International Statistical Review*, 69(1), 11-15.
- Nijdam, B (2003). *Statistiek in onderzoek 1: Beschrijvende technieken*. Groningen: Wolters-Noordhoff.
- Nijdam, B. (2004). *Statistiek in onderzoek 2: Inductieve technieken*. Groningen: Wolters-Noordhoff.
- Nijdam, B., & Van Buuren, H. (1999). *Statistiek voor de sociale wetenschappen 1. Beschrijvende statistiek*. Groningen: Wolters-Noordhoff.
- Nijdam, B., & Van Buuren, H. (1999). *Statistiek voor de sociale wetenschappen 2. Inleiding in de inductieve statistiek*. Groningen: Wolters-Noordhoff.
- Noll, J. (2007). *Statistics teaching assistants' statistical content knowledge and pedagogical content knowledge*. Unpublished Unpublished doctoral dissertation, Portland State University., Portland, OR.
- Ohlsson, S. (1996). Learning to Do and Learning to Understand: A Lesson and a Challenge for Cognitive Modeling. In P. Reimann & H. Spada (Eds.), *Learning in Humans and Machines* (pp. 37-62). Pergamon.
- Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics Anxiety: nature, etiology, antecedents, effects, and treatments--a comprehensive review of the literature. *Teaching in Higher Education*, 8(2), 195-209.
- Oost, H., & De Jong, J. (1996). Cursus 'Scriptie schrijven'. In J. Kaldeway, J. Haenen, S. Wils & G. Westhoff (Eds.), *Leren leren in didactisch perspectief* (pp. 214-231). Groningen: Wolters-Noordhoff.
- Osana, H.P. & Seymour, J.R. (2004). Critical Thinking in Preservice Teachers: A Rubric for Evaluating Argumentation and Statistical Reasoning. *Educational Research and Evaluation*, 10(4-6), 473-498



- Panizzon, D., & Pegg, J. (1997). *Investigating Students' Understandings of Diffusion and Osmosis: A Post Piagetian Analysis*. Paper presented at the 1997 annual conference of the Australian Association for Research in Education, December, 1997, Brisbane.
- Panizzon, D., Arthur, D., & Pegg, J. (1997). ESSA: Developing and scoping a test to explore scientific literacy and achievement in NSW. *Teaching Science*, 52(4), 22-27.
- Pegg, J. (1997). Broadening the Descriptors of van Hiele's Levels 2 and 3. in F. Biddulph & K. Carr (Eds), *People in Mathematics Education* (pp. 391-396)., Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia, July, 1997, Rotorua.
- Perneger, T.V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, 316, 1236-1238.
- Perney, J., & Ravid, R. (1991). *The Relationship Between Attitudes Towards Statistics, Math Self-Concept, Test Anxiety and Graduate Students' Achievement in an Introductory Statistics Course*. Evanston, IL: National College of Education.
- Pfannkuch, M. (1999). *Characteristics of Statistical Thinking in Empirical Inquiry*. Unpublished Unpublished Doctoral Thesis, The University of Auckland.
- Pieters, J.M., & Verschaffel, L. (2003). Beïnvloeden van leerprocessen. In N. Verloop & J. Lowyck (red.), *Onderwijskunde* (251-293). Groningen: Wolters-Noordhoff.
- Pintrich, P.R. (1994). Continuities and discontinuities: future directions for research in educational psychology. *Educational Psychologist*, 29(3), 137-148.
- Pintrich, P.R. (2002). The role of metacognitive knowledge in learning, teaching and assessing. *Theory Into Practice*, 41(4), 219-225.
- Pintrich, P.R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.
- Pintrich, P.R., & De Groot, E. V. (1990). Motivation and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-40.
- Pintrich, P.R., & Schunk, D.H. (2002). *Motivation in Education* (2nd. ed.). New Jersey: Merrill: Prentice Hall.
- Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor - Michigan: University of Michigan.
- Pintrich, P.R., Smith, D.A. F., Garcia, T., & McKeachie, W.J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire. *Educational and Psychological Measurement*, 53, 801-813.
- Quilici, J.L., & Mayer, R.E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, 16(3), 325-342.
- Ramsden, P. (1992). *Learning to Teach in Higher Education*. London: Routledge.
- Reading, C. (1996). *An investigation into Students' Understanding of Statistics*. Unpublished Unpublished Doctoral Dissertation, University of New England, Armidale.
- Reading, C. (1999). *Understanding Data Tabulation and Representation*. Paper presented at the 23rd Conference of the International Group for the Psychology of Mathematics Education, Haifa, Israel.
- Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal*, 3(2), 87-105.
- Reigeluth C.M. (1992) Reflections on the Implications of Constructivism for Educational Technology. In T.M. Duffy & D.H. Jonassen (Eds.), *Constructivism and the Technology of Instruction: A conversation*. Lawrence Erlbaum Associates.



- Roorda, G., & Van Streun, A. (2002). Criteria voor samenhangend (wiskunde)onderwijs. Een oriënterend onderzoek naar aanwijzingen voor samenhangend onderwijs. In J. F. Deinum, J. Van Maanen, A. Van Streun & J. Tolboom (Eds.), *Werken aan de kwaliteit van het onderwijs in de bètavakken*. Groningen: Universiteit van Groningen.
- Rosenthal, B. (1992). No More Sadistics, No More Sadists, No More Victims. *UMAP Journal*, 13, 281-290.
- Royalty, G.M., Gelso, C.J., Mallinckrodt, B., & Garrett, K. (1986). The environment and the student in counseling psychology: Does the research training environment influence graduate students' attitudes toward research? *The Counseling Psychologist*, 14, 9-30.
- Ryan, A.M., & Deci, E.L. (2000). The darker and brighter sides of human existence: basic psychological needs as a unifying concept. *Psychological Inquiry*, 11(4), 319-338.
- Ryan, A.M., & Pintrich, P.R. (1997). Should I ask for help? The role of motivation and attitudes in adolescents' help-seeking in math class. *Journal of Educational Psychology*, 89, 329-341.
- Ryan, R.M., & Deci, E.L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, 25(54-67).
- Salsburg, D.S. (1985). The Religion of Statistics as Practiced in Medical Journals *The American Statistician*, 39 (3), 220-223.
- Sarfo, F.K. (2005). *Developing Technical Expertise in Secondary Technical Schools in Ghana: The Effect of Powerful Learning Environments With and Without ICT and the Moderating Effect of Instructional Conceptions*. Leuven: Katholieke Universiteit Leuven.
- Savitz, D.A., & Olshan, A.F. (1995). Multiple Comparisons and related Issues in the Interpretation of Epidemiologic Data. *American Journal of Epidemiology*, 142(9), 904-908.
- Schau, C.G., Dauphinee, T.L., & Del Vecchio, A. (1992). The development of the survey of attitudes toward statistics. San Francisco: American Educational Research Association.
- Schau, C.G., & Mattern, N. (1997). Use of map techniques in teaching statistics courses. *The American Statistician*, 51(2), 171-175.
- Schau, C.G., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, 55(5), 868-875.
- Schaufeli, W., Bakker, A., & De Jonge, J. (2003). *De Psychologie van Arbeid en Gezondheid*. Houten: Bohn Stafleu Van Loghum.
- Schmidt, H.G. (1997). Drie factoren die verhinderen dat mensen zelfstandig leren kennis verwerven. *Tijdschrift voor Hoger Onderwijs*, 15(1), 26-33.
- Scholten, I., Keeves, J.P., & Lawson, M.J. (2022). Validation of a free response test of deep learning about the normal swallowing process. *Higher Education*, 44, 233-255.
- Schouwenburg, H. C. (1998). Invloeden op studieresultaat. Retrieved August, 20, 2004, from <http://www-dsz.service.rug.nl/bss/so/topics/research/lee.htm>
- Schunk, D.H. (1982a). Effects of effort attributional feedback on children's perceived self-efficacy and achievement. *Journal of Educational Psychology*, 75, 848-856.
- Schunk, D.H. (1982b). Verbal self-regulation as a facilitator of children's achievement and self-efficacy. *Human Learning*, 1, 265-277.
- Schunk, D.H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review*, 1, 173-208.
- Schunk, D.H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26, 207-231.
- Schunk, D.H. (2005). Self-regulated learning: the educational legacy of Paul R. Pintrich. *Educational Psychologist*, 40(2), 85-94.

- Schuyten, G. (2001). Discussion. Research Skills: a Closely Connected Triplet of Research Area, Research Methodology and Statistics. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 227-230). Voorburg: International Association for Statistical Education and International Statistical Institute.
- Shaughnessy, J.M. (1992). Research in probability and statistics: reflections and directions. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York NY: Macmillan.
- Shuell, T.J. (1988). The role of the student in learning from instruction, *Contemporary Educational Psychology*, 13, 276-295.
- Sijtsma, K., & Van Duijn, M.A.J. (1999). Computerondersteund statistiekonderwijs aan de Nederlandse universiteiten: de stand van zaken. *Kwantitatieve methoden*, 20(60), 5-7.
- Singley, M.K. & Anderson, J.R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Skemp, R.R. (1971). *The Psychology of Learning Mathematics*. Harmondsworth, Middlesex: Penguin.
- Sorge, C., & Schau, C.G. (2002). *Impact of engineering students' attitudes on achievement in statistics: a structural model*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Spector, P.E. (1992). *Summated Rating Scale Construction*. Newbury Park, CA: Sage.
- Spence, J.T., & Helmreich, R.L. (1983). Achievement-Related Motives and Behaviors. In J.T. Spence (Ed.), *Achievement and Achievement Motives* (pp. 7-74). San Francisco: W.H. Freeman and Company.
- Stoop, I.A.L. (2005). *De jacht op de laatste respondent*. Utrecht: Universiteit Utrecht.
- Tabachnick, B.G. & Fidell, L.S. (2001). *Using multivariate statistics* (4<sup>th</sup> ed.). Boston: Allyn and Bacon.
- Tang, C. (1998) Effects of collaborative learning on the quality of assignments. In B. Dart & G. Boulton-Lewis (Eds), *Teaching and Learning in Higher Education* (pp. 102-123). Melbourne: ACER.
- Tempelaar, D. (2002). *Modelling students' learning of introductory statistics*. Paper presented at the ICOTS6, Cape Town, South Africa.
- Tempelaar, D.T. (2004, August, 1-4). *Statistical reasoning assessment: an analysis of the SRA instrument*. Paper presented at the ARTIST Roundtable Conference on Assessment in Statistics, Lawrence University.
- Tempelaar, D.T. (2007). *Expectancy-value based achievement motivations and their role in student learning*. Maastricht: Universiteit van Maastricht.
- Ter Laak, J.J.F., & De Goede, M.P.M. (2003). *Psychologische diagnostiek. Inhoudelijke en methodologische grondslagen*. Lisse: Swets & Zeitlinger.
- Tomic, W., & Span, P. (Eds.). (1993). *Onderwijspsychologie: beïnvloeding, verloop en resultaten van leerprocessen*. Utrecht: Lemma.
- Trigwell, K. (1999) Using evidence of outcomes to evaluate assessment practices, in: C. Rust (Ed.) *Improving Student Learning: Outcomes*, 359-366. Oxford: OCSLD.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Ullman, N.R. (1996). Statistical or quantitative thinking - a fundamental intelligence. *Journal of Statistical Education* [on-line], 2(1).
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74-79.

- Van Buuren, H., & Giesbertz, W. (1999). Naar een virtueel onderzoekcentrum. Innovatie van het methoden en technieken van onderzoeksonderwijs. In K. Schlusmans, R. Slotman, C. Nagtegaal & G. Kinkhorst (Eds.), *Competentiegerichte leeromgevingen* (pp. 137-150). Utrecht: Lemma.
- Van Buuren, H., & Hummel, H. (1997). *Onderzoek de Basis*. Groningen: Wolters-Noordhoff.
- Van Buuren, J.A., Bijker, M.M., & Wynants, G. (2005). *Instrumenten voor de effectmeting van competentiegericht onderwijs bij psychologiestudenten in afstandsonderwijs*. Paper presented at the 32ste Onderwijs Research Dagen, Gent, België.
- Van Buuren, J.A., & Giesbertz, W.J. (1998). *Innovatie M&T Onderwijs. Naar een Virtueel Onderzoek Centrum*. Heerlen: Open Universiteit Nederland.
- Van den Akker, J.J.H. (1998). *De uitbeelding van het curriculum*. Enschede: Universiteit Twente.
- Van den Bercken, J.H.L. & Voeten, M.J.M. (2002). *Variantieanalyse*. Groningen: Wolters-Noordhoff.
- Van der Klink, M.R., & Boon, J. (2003). Competencies: the triumph of a fuzzy concept. *International Journal of Human Resources Development and Management*, 3(2), 125-137.
- Van der Sanden, J.M.M. (2004). *Ergens goed in worden. Naar leerzame loopbanen in het beroepsonderwijs*. Eindhoven: Fontys Pedagogisch Technische Hogeschool.
- Van Merriënboer, J.J.G. (1997). *Training Complex Cognitive Skills: a Four-Component Instructional Design Model for Technical Training*. New Jersey: Englewood Cliffs.
- Van Merriënboer, J.J.G. & Dijkstra, S. (1997). The Four-Component Instructional Design model for training complex cognitive skills. In R.D. Tennyson, F. Schott, N. Seel, & S. Dijkstra (Eds.), *Theory, research, and models of instructional design: international perspective* (Vol. 1, pp. 427-447). Mahwah, NJ: Lawrence Erlbaum.
- Van Merriënboer, J.J.G. & Kanselaar, G. (2006). Waar staan we na 25 jaar onbewijstechnologie in Vlaanderen, Nederland en de rest van de wereld? *Pedagogische Studiën*, 83, 278-300.
- Van Merriënboer, J.J.G. & Kirschner, P.A. (2007). *Ten Steps to Complex Learning. A Systematic Approach to Four-Component Instructional design*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Van Merriënboer, J.J.G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147-177.
- Vansteenkiste, M., Simons, J., & Lens, W. (2004a). Motivating learning, performance, and persistence: the synergetic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology*, 87(2), 246-260.
- Vansteenkiste, M., Simons, J., Lens, W., & Soenens, B. (2004b). De kwaliteit van motivatie telt: over het promoten van intrinsieke doelen op een autonomie ondersteunende wijze. *Nederlands tijdschrift voor de psychologie*, 59, 119-131.
- Veldhuis-Diermanse, A.E. (2002). *CSC Learning? Participation, learning activities and knowledge construction in computer-supported collaborative learning in higher education*. Wageningen: Wageningen Universiteit.
- Verloop, N., & Lowyck, J. (Eds.). (2003). *Onderwijskunde*. Groningen: Wolters-Noordhoff.
- Verschuren, P.J.M. (1996). Een adequate probleemstelling voor sociaal-wetenschappelijk praktijkgericht onderzoek. *Sociologische Gids*, XLIII(4), 280-290.
- Verschuren, P.J.M. (2002). Falende universiteit. Voer voor methodologen. *Sociologische Gids*, 49(4), 349-370.
- Verschuren, P.J.M. (2002). *Dogma's en ontwikkelingen in wetenschap en methodologie: bedreigingen en kansen*. Nijmegen/Wageningen. Oratie.
- Verschuren, P.J.M. (2003). Profielen en competenties in (kwalitatief) onderzoek. *Kwalon*, 8(3), 21-24.
- Visitatiecommissie Psychologie (2001). *Onderwijsvisitatie Psychologie*. Utrecht: VSNU.

- Vrugt, A. (1994). Waargenomen eigen competentie, sociale vergelijking, affectieve reacties en studieprestaties. *Nederlands tijdschrift voor de psychologie*, 49, 171-177.
- Ward, B. (2004). The best of both worlds: A hybrid statistics course. *Journal of Statistics Education*, 12(3), 74-79.
- Watson, J.M., & Callingham, R.A. (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Journal*, 2(2), 3-46.
- Watson, J.M., & Moritz, J.B. (2000). Development of understanding of sampling for statistical literacy. *The Journal of Mathematical Behavior*, 19, 109-136.
- Wild, C.J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, 67(3), 223-248.
- Wild, K.P., & Schiefele, U. (1993). Induktiv versus deduktiv entwickelte Fragebogenverfahren zur Erfassung von Merkmalen des Lernverhaltens. *Unterrichtswissenschaft*, 21, 312-326.
- Willems, J. (1981). Problem-based (group) teaching: A cognitive science approach to using available knowledge. *Instructional Science*, 10, 5-21.
- Willemsen, E.W. & Gainen, J. (1995). Reenvioning Statistics: A Cognitive Apprenticeship Approach. *New Directions for Teaching and Learning*, 61, 99-108.
- Wisembaker, J.M., Scott, J.S., & Nasser, F.M. (1999). A cross-cultural comparison of path models relating attitudes about and achievement in introductory statistics courses. Paper presented at the 52nd Session of the International Statistical Institute, Helsinki, Finland.
- Wisembaker, J.M., Scott, J.S., & Nasser, F.M. (2002). Structural equation models relating attitudes about and achievement in introductory statistics courses: a comparison of results from the U.S. and Israel. [On-line].
- Wolfe, E.W. (2000). Equation and Item Banking with the Rasch Model. *Journal of Applied Measurement*, 1(4), 409-434.
- Wolfe, E.W. (2005). Making Measures. *Journal of Applied Measurement*, 6(2), 245-246.
- Wolfe, E.W., & Chiu, C.W.T. (1997). Measuring change over time with a Rasch rating scale model. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, USA., Chicago, Illinois
- Wouters, P. (2007). *How to optimize cognitive load for learning from animated models*. Heerlen: Open Universiteit Nederland.
- Wright, B.D., & Bell, S.R. (1984). Item banks: What, why, how. *Journal of Educational Measurement*, 2, 331-345.
- Wright, B.D., & Masters, G.N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.
- Wynne, H., Lockhorst, D., & Westers, P. (1999). Het COSO-Project: Computer Ondersteund Statistiek Onderwijs. *Kwantitatieve Methoden*, 60, 29-40.
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum.
- Zieffler, S.A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Unpublished Ph.D. Thesis, University of Minnesota.
- Zusho, A., & Pintrich, P.R. (2003). Skill and will: the role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education*, 25(9), 1081-1094.



## Samenvatting

In dit proefschrift wordt beschreven hoe de vakken statistiek en methoden en technieken van onderzoek (M&T) in de vorm van competentiegericht onderwijs zijn herontworpen en geïmplementeerd in het curriculum van de opleiding psychologie van de Open Universiteit Nederland. Door afstand te nemen van de autonomie van de traditionele M&T en statistiekvakken en deze te integreren tot een coherent geheel met psychologische leerinhouden wordt een betere afstemming met het streefdoel van universitair onderwijs gerealiseerd: de zelfstandige bestudering van de wetenschap door middel van (psychologisch) onderzoek.

Hoofdstuk 1 beschrijft de aanleiding voor dit proefschrift. De literatuur verwijst naar drie mogelijke oorzaken voor de problemen die studenten hebben met het leren, zich herinneren en toepassen van statistiek. De eerste verklaring zoekt de oorzaak bij de studenten zelf: een deficiënte vooropleiding, een negatieve attitude, angst, geringe motivatie en weinig zelfvertrouwen. De tweede verklaring grijpt terug op de inhoud van het vak statistiek zelf. Deze is abstract en complex en wordt in compacte cursussen voornamelijk in de beginfase van de studie gegeven, op het moment dat studenten nog niet beseffen wat onderzoek inhoudt. De derde verklaring zoekt de oorzaak in de positie die het vak statistiek in het curriculum heeft gekregen.

Een oplossing voor deze problemen wordt in dit proefschrift gezocht in het herontwerp van het statistiekonderwijs in de vorm van een competentiegerichte leeromgeving. Het idee, dat ten grondslag ligt aan het herontwerp is om de student geïntegreerde hele onderzoektaken in een doorlopende leer- en toepassingslijn aan te bieden, waardoor de student steeds weer wordt uitgedaagd om de hele onderzoekscyclus te doorlopen en een variëteit aan verschillende psychologische onderzoekstaken uit te voeren. De verwachting is dat de student in deze leeromgeving op een natuurlijke wijze vertrouwd raakt met het toepassen van statistiek in een verscheidenheid aan psychologische contexten, de bijbehorende juiste methoden leert gebruiken en de verschillende bijpassende technieken weet te selecteren. De onderzoekstaken beogen tevens een bijdrage te leveren aan het beter doorgronden van de betekenis van statistiek voor psychologisch onderzoek, aan een verbeterde motivatie voor het doen van onderzoek en aan een meer positieve attitude ten aanzien van statistiek. Op basis van deze verwachtingen worden de onderzoeksvragen geconcretiseerd. Voor de reconstructie van het ontwerpproces wordt gebruik gemaakt van Intervention Mapping. Dit is een systematiek voor planmatige veranderingen.

Hoofdstuk 2 beschrijft in detail drie stappen die volgens Intervention Mapping na de probleemanalyse moeten worden gezet om zicht te krijgen op wat er moet veranderen om de problemen op te lossen. De eerste stap is de uitwerking van de gewenste gedragingen ('performance objectives' of gedragsdoelen). De tweede stap is het achterhalen van belangrijke en veranderbare determinanten die deze gedragingen beïnvloeden.

Aangezien de positionering van het vak statistiek in het psychologiecurriculum als dé oorzaak van de problemen met het statistiekonderwijs wordt beschouwd, worden de determinanten gezocht in de (externe) leeromgeving. Deze bepaalt voor een groot gedeelte hoe docenten hun onderwijs vorm kunnen geven. De determinanten bepalen echter niet alleen de leeruitkomsten, maar ook de wijzen waarop studenten leren (inzet van leerstrategieën) die weer afhankelijk is van hoe betekenisvol, waardevol, nuttig en veilig ze het onderwijs vinden (motivatie, attitude). Dus met veranderingen in de positionering van het vak statistiek in het curriculum wordt ook beoogd de in hoofdstuk 1 vermelde motivationele en attitudeproblemen op te lossen. Als laatste stap wordt in hoofdstuk 2 de determinanten afgestemd op gedragsdoelen. De hieruit afgeleide veranderdoelen worden in een matrix afgebeeld. Hiermee is het interventiedoel geconcretiseerd.

Hoofdstuk 3 richt zich op de onderwijskundige richtlijnen die de veranderdoelen vorm kunnen geven. Richtlijnen om het abstracte en complexe karakter van de statistiekinhoud voor studenten studeerbaar te maken en die toepasbaar zijn bij het uitvoeren van psychologisch onderzoek zijn gezocht binnen de 'Cognitive Load Theory' (Van Merriënboer & Sweller, 2005), 'Meaningful Learning' (Ausubel, 1960, 1968), het 'Cognitive Apprenticeship Model' (Collins, Brown, & Newman, 1989) en de Theory of the Research Environment (Gelso, 2006). De meest belangrijke richtlijnen zijn: het activeren van voorkennis, het gebruik van advance organizers, gelegenheid bieden tot oefenen en toepassen in een langlopende leer- en toepassingslijn, het voorkomen van cognitieve overbelasting, en het gebruik van een variëteit van authentieke onderzoeksvraagstukken.

In hoofdstuk 4 wordt, op basis van de bevindingen uit hoofdstuk 2 en 3, uiteengezet hoe de leertheoretische principes en onderwijskundige richtlijnen zijn toegepast in het herontwerp van de leeromgeving. Het ontwerp sluit goed aan bij het gedachtengoed van het 4C/ID-model (Van Merriënboer, 1997), een model dat zich richt op de verwerving van complexe vaardigheden. Bij het herontwerp zijn twee statistiekcursussen, een SPSS-cursus, drie M&T-cursussen betrokken. De bachelorthesis wordt als eindniveau van de onderzoekscompetentie beschouwd. In zes opeenvolgende onderzoekspractica (net zoveel als de geïntegreerde cursussen) dienen de studenten zich de herverkavelde leerstofinhouden van statistiek, M&T en psychologie eigen te maken. Het ontwerp van de practica is in twee pilots getest. De bevindingen van beide pilots wijzen op de bruikbaarheid van dit herontwerp van het statistiekonderwijs. De integratie van statistiek, M&T en psychologie wordt positief gewaardeerd, men lijkt de angst voor statistiek te overwinnen, de studielast blijkt haalbaar en de elektronische leeromgeving blijkt uit te nodigen tot kritische discussies en uitwisselingen over de onderzoeksbevindingen tussen studenten onderling. Besloten is het ontwerp te implementeren in het psychologiecurriculum onder de naam onderzoekscompetentie.

Wat niet uit de pilots kan worden opgemaakt, is of studenten ook betere leerresultaten behalen dan in de conventionele aanpak van het statistiekonderwijs.

Hoofdstuk 5 beschrijft een crosssectioneel onderzoek dat is uitgevoerd onder studenten die het vakgerichte statistiekonderwijs hebben gevolgd en studenten die het allereerste onderzoekspracticum – waarin studenten voor het eerst kennismaken met



correlatie, regressie en variantieanalyse – hebben doorlopen. Verondersteld werd dat studenten in de competentiegerichte, herontworpen leeromgeving een positievere attitude hebben ten aanzien van statistiek, beter gemotiveerd zijn om psychologisch onderzoek uit te voeren, meer adequate leerstrategieën hanteren en over meer kennis en vaardigheid beschikken op het gebied van statistiek dan in de traditionele, vakgerichte leeromgeving. De onderzoeksresultaten bevestigen de hypothesen: de competentiegerichte leeromgeving onderscheidt zich in gunstige zin van de traditionele leeromgeving ten aanzien van alle genoemde factoren, hoewel de effectgrootten niet groot zijn. Wel leidt de analyse tot nieuwe inzichten van de werking van attitudes en motivatie. De tendensen wijzen erop dat in de vakgerichte benadering van het statistiekonderwijs de (negatieve) attitude ten aanzien van de statistiek de overhand heeft en dat de motivatie nauwelijks wordt opgewekt, terwijl in de competentiegerichte benadering de (negatieve) attitude geneutraliseerd wordt en dat er meer gemotiveerd en diepgaand wordt gestudeerd.

Hoofdstuk 6 behandelt de opzet en resultaten van een longitudinaal onderzoek dat is uitgevoerd onder een groep studenten uit het competentiegerichte statistiekonderwijs en een groep studenten uit het vakgerichte statistiekonderwijs. De voormeting vond drie jaar eerder plaats onder studenten die in een van de pilots hadden geparticipeerd en studenten die in dezelfde periode met een cursus M&T begonnen. Drie jaar later zijn deze studenten opnieuw bevraagd met de vragenlijst die ook in het crosssectionele onderzoek is gebruikt. Ook nu weer werd het duidelijk dat de competentiegerichte benadering zich weer gunstig onderscheidde van de traditionele leeromgeving. Er bleken zelfs significante interactie-effecten van de leeromgeving in de tijd op te treden. Wederom zijn de effectgrootten bescheiden maar ze blijken duurzaam.

Hoofdstuk 7 reflecteert op de bevindingen en resultaten uit de eerdere hoofdstukken en de implicaties daarvan voor het statistiekonderwijs, niet alleen voor de studenten psychologie van de Open Universiteit Nederland maar tevens voor het statistiekonderwijs in meer algemene zin. Gesteld mag worden dat door de integratie met M&T en psychologie het vak statistiek nu duidelijker dan voorheen een meer verborgen (want geïntegreerde) én een meer prominente (want overal aanwezige) plaats in het curriculum gekregen. De opleiding Psychologie van de Open Universiteit Nederland heeft daarmee een eerste stap gezet om de onderzoekscompetentie vorm te geven. Integratie van statistiek, M&T en psychologie is tegelijk een uitdagende stap. Integratie opent het perspectief voor nauwe samenwerking en afstemming tussen docenten van verschillende disciplines en voor meer begrip voor elkaars vak.





## Summary

This thesis describes how statistics education has been redesigned and implemented in the curriculum of the school of psychology of the Open Universiteit Nederland. In stead of autonomous and separate traditional statistics service courses and courses in research methods, both kinds of courses have been integrated into a coherent whole, including psychological theory, to realize a better fit with the educational target goal: the autonomous study of science by carrying out (psychological) research.

Chapter 1 describes the arguments for this thesis. The literature refers to three possible causes for the problems that university students have in learning, remembering and using statistics. The first explanation looks at causes which can be ascribed to the students themselves: a deficient preparation in mathematics, a negative attitude, anxiety, low motivation and low self-esteem. The second type of explanation goes back to the contents of the subject of statistics. This subject is abstract and complex, and is taught in a compressed way mainly in the initial phase of the study, at a time when students are not aware of the meaning of research. The third explanation is found in the position of statistics within the curriculum.

In this thesis a solution for these problems has been provided by redesigning statistics education into a competence-based learning environment. The underlying idea for redesigning is to create a long-term programme for students in which a series of whole-psychological-research-tasks is carried out in a cumulative way. The expectation is that students in such a learning environment naturally become familiar with selecting and applying appropriate statistics techniques in a variety of psychological research contexts. Additionally the research assignments intend to support a better understanding of the value and importance of statistics in psychological research, to improve motivation for carrying out research and to establish more positive attitudes towards statistics. Based on these expectations, concrete research questions have been formulated. For the (reconstruction of the) design process the Intervention Mapping approach has been used. The purpose of Intervention Mapping is to provide program planners with a framework for effective decision making at each step in intervention planning, implementation, and evaluation.

Following the needs assessment chapter 2 describes the first three steps in Intervention Mapping which provide insight in what needs to be changed to solve the problems. The first step consists of a detailed description of the desired behaviours (performance objectives). The second step is to identify important and variable determinants that precede and influence the desired behaviours. Since the posi-

tioning of statistics in the psychology curriculum is considered to be the cause of the problem in statistics education, determinants are investigated in the (external) learning environment. The discipline's position determines to a large extent how teachers can shape their education. The determinants not only define students' learning outcomes, but also students' learning strategies which in turn depend on how meaningful, valuable, useful, and secure students think education is (motivation, attitude). As a consequence, changing statistics education's position in the psychology curriculum, as mentioned in chapter 1, most probably will lead to the desired changes in motivations and attitudes and to a structural solution of recognized problems in statistics education. Finally, in chapter 2 determinants are tuned with the performance objectives. The resulting change objectives are pictured in a matrix. By doing so the goal of the intervention has been made concrete.

Chapter 3 focuses on the educational guidelines which will shape the change objectives.

Guidelines which may scaffold and support the learning processes which are needed for the abstract and complex statistics content and which are applicable in carrying out psychological research have been found in "Cognitive Load Theory" (Van Merriënboer & Sweller, 2005), 'Meaningful Learning' (Ausubel, 1960, 1968), the 'Cognitive Apprenticeship Model' (Collins, Brown, & Newman, 1989) and the 'Theory of the Research Environment' (Gelso, 2006). The most important guidelines are: the activation of prior knowledge, the use of advance organizers, the provision of opportunities to deliberately practise and to apply statistics in a long-term programme of learning and application, the prevention of cognitive overload, and the design of a variety of authentic psychological research tasks.

Chapter 4 explains, based on the findings from chapter 2 and 3, how the learning and instructional principles and guidelines have been applied in the redesign of the learning environment. The design adequately fits the ideas of the 4C/ID-model (Van Merriënboer 1997), an instructional model that emphasizes the acquisition of complex skills. The redesign of the curriculum concerns two statistics service courses statistics, a SPSS-course, and three research methods courses. The bachelor thesis has been considered as the final stage of the research competency. In six consecutive research practicals (just as much as the number of integrated courses) the students become familiar with the reallocated statistics contents, research methods and psychology. The design of the integrated research practicals has been tested in two pilots. The formative evaluation of the redesign confirms its usefulness for statistics education for future psychologists. The integration of statistics, research methods, and psychology is appreciated positively, seems to help students overcoming their statistics anxiety, the studyload appears feasible and the electronic learning environment appears to trigger active engagement, exchanges of messages and critical discussions about findings between students. It has been decided to implement the statistics education design into the psychology curriculum using the label 'research competency'.

The pilots did not offer the opportunity to investigate whether students achieved better learning outcomes than in the conventional approach of statistics education. Chapter 5 describes a cross-sectional investigation that was conducted among students who have been taught statistics education in the subject-oriented or traditional way and students who have carried out the first research practical, in which students meet for the first time statistics techniques like correlation, regression and variance-analysis. It is supposed that students in the redesigned competence-based learning environment gain a more positive learning attitude towards statistics, become motivated to conduct psychological research, will use more appropriate learning strategies, will gain more insight in statistics and better established statistics skills than it is the case in the traditional, subject-oriented learning environment. The research results confirm the assumptions: the competence-based learning environment distinguishes itself in a favourable sense in comparison with the traditional learning environment with regard to all these factors, although the effect-sizes are not large. Notable are the mechanisms of attitude and motivation. It seems that in the subject-oriented approach to statistics education the (negative) attitude towards statistics prevails and that the motivation is hardly triggered, while in the competence-based approach the (negative) attitude has been neutralized and that students use a motivated and deep approach to study statistics.

Chapter 6 deals with results of a longitudinal study that was carried out among a group of students from the competence-based statistics education and a group of students from the subject-oriented statistics education. The pre-measurement took place three years earlier among students in one of the pilots and students who had participated in the same period in a research methods course, meaning that this group of students already participated in one or more statistics service courses. Three years later, both groups of students have been studied for the second time using the questionnaire which also has been used in the cross-sectional research. Once again it became clear that the competence-based approach is favourably different from the traditional learning environment. Apparently the learning environment even induced significant interactions over time and again effect sizes were significant, though not impressive.

Chapter 7 reflects on the findings and results from the previous chapters and their implications for statistics education not only for psychology students of the Open Universiteit Nederland but also for statistics education in a more general sense. It may be inferred that by integrating statistics and research methods into authentic psychological and professional learning tasks statistics has obtained both a more (psychologically) hidden (integrated) and a more prominent (omnipresent) position within the curriculum.

The school of Psychology of the Open Universiteit Nederland has taken the first step in shaping the research competency. Integration of statistics, research methods, and psychology is a challenging step. Integration opens the perspective for

shared goals, close cooperation and coordination between teachers of different disciplines and for a better understanding of each other's profession.

## Dankwoord

De omslag van het vakgerichte naar het competentiegerichte statistiekonderwijs en de daarmee gepaard gaande uitwerkingen van ideeën, tests en pilots is een creatief proces geweest. Het voorliggende proefschrift is een verslag van dit proces dat een tienjarige periode van innoveren, experimenteren en veranderen omvat. Aan mijn promotores komt de eer toe dat ze me ervan hebben weten te overtuigen het verslag daarvan beknopt en 'to the point' te houden. Jasper von Grumbkow en Marcel van der Klink veel dank hiervoor.

Dank ook aan Henk van der Molen en René Hezewijk, successieve decanen van de faculteit Psychologie voor de gelegenheid die zij mij hebben geboden het onderwijs te herontwerpen in de door mij beoogde richting. De onderzoekscompetentie wordt inmiddels door een grote groep collegae gedragen.

Een woord van dank mag zeker uitgaan naar degenen in mijn omgeving die delen van het manuscript zo consciëntieus hebben becommentarieerd, zoals Esther Bakker, Francine Dehue, Henk Dubbelt, Tilly Houtmans, Aart Mudde, Olga Firssova, Frits Kluijtmans, Mienieke Pouwels, Jef Syroit en Iwan Wopereis. Een bijzonder woord van dank aan Karel Kreijns die het mogelijk heeft gemaakt dat vele studenten de vragenlijsten digitaal konden verzenden.

Van de studenten die door de nieuwe onderwijsbenadering zeer enthousiast zijn geraakt om zelf ook onderzoek te gaan doen en die samen met mij de empirische evidentie van het nieuwe curriculum hebben verkend, noem ik in de eerste plaats Monique Bijker die nu haar 'flow' heeft gevonden in sociaalwetenschappelijk onderzoek. In de tweede plaats Gabriëlle Wynants die de diverse onderwerpen heeft aangescherpt. Daarnaast dank ik de doctoraalstudenten die over delen van het onderzoek naar de effecten van competentiegericht onderwijs hun scriptie hebben geschreven. Veel dank ben ik ook verschuldigd aan de studenten van de faculteit Psychologie die kostbare tijd hebben besteed om de lange vragenlijsten in te vullen. Wat ik hen hoop terug te geven is goed ontworpen en betekenisvol onderwijs.

De fraaie omslag en het binnenwerk van mijn proefschrift is verzorgd door Annette Bouwels en Sandra Daems, een bijzonder woord van dank hiervoor. Dank ook aan Jan Hendriks voor de redactionele correcties. Monique Verschuren was mijn steun en toeverlaat bij alle voorkomende klussen.

Een woord van dank ook aan mijn voormalige collega Daantje Derks die met een 'kwink en een wink ☺' steeds weer in staat bleek om de gezelligheid in stand te houden. Een bijzondere plaats wordt ingenomen door mijn paranimfen en 'brothers in arms' Bert Nijdam en Wil Giesbertz. Bert deelde mijn belangstelling voor het structureren van het statistiekonderwijs vanuit onderzoeksvragen. Met Wil maakte ik de Gestaltswitch naar het denken buiten de grenzen van het vak statis-

tiel. Samen maakten we de stap naar het competentiegerichte onderwijsontwerp. Niet in de laatste plaats gaat een woord van dank naar mijn gezin: Nel, mijn sparringpartner, die steeds de uitdaging aanging om op indringende wijze met mij de kritische discours aan te gaan en de didactisch-pedagogische disciplines vanuit verschillende perspectieven benaderde met het accent op de beleving van de student zelf. Je hebt me alert gehouden en mijn kritische denken gestimuleerd. Vanzelfsprekend dank ik ook Paul, Leonie en Charlotte die hun nauwe betrokkenheid toonden tijdens het bijwonen van de 'international conferences on teaching statistics' in Kaapstad en Salvador de Bahia (en mij na afloop precies wisten te vertellen wat nog allemaal voor verbetering vatbaar was).

Met dit proefschrift zet ik een nieuwe koers uit. Het onderwerp vraagt erom om verder uitgedragen te worden. De komende periode zullen er dan ook artikelen verschijnen voor de internationale gemeenschap.

Heerlen, 21 april 2008

Hans van Buuren

## Lijst van eigen publicaties en presentaties in relatie tot het thema van dit proefschrift

- Bijker, M.M., & Van Buuren, J.A. (2006). *Effectmeting van competentiegericht onderwijs bij psychologiestudenten in afstandsonderwijs. Effecten op motivaties, attitudes, leerstrategieën en kennis*. Paper presented at the 33ste Onderwijs Research Dagen, Amsterdam, Nederland.
- Bijker, M.M., Wynants, G., & Van Buuren, J.A. (2006). *A comparative study of the effects of motivational and attitudinal factors on studying statistics*. Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Bahia, Brazilië.
- Giesbertz, W., & Van Buuren, H. (2004). Competentiebegrip bij de OUNL. Van hype naar realistische praktijk. In L. Dercksen, H. Dekker, M. van der Klink & S. Wagenaar (Eds.), *Competentiegericht leren: de context centraal* (Vol. 5(3), pp. 80-88). Doetinchem: Reed Business Infomation.
- Hutjes, J.M., & Van Buuren, J.A. (1992). *De gevalstudie. Strategie van kwalitatief onderzoek*. Meppel/Amsterdam: Boom.
- Joosten-ten Brinke, D., Verhoeven, P.S., & Van Buuren, J.A. (2003). Peerassessment en selfassessment in een competentiegericht methodologiecurriculum. *Tijdschrift voor Hoger Onderwijs*, 21(4), 273-286.
- Kreijns, K., Kirschner, P.A., Jochems, W., & Van Buuren, H. (2004). Determining Sociability, Social Space, and Social Presence in (A)synchronous Collaborative Groups. *CyberPsychology & Behavior*, 7(2), 155-172.
- Kreijns, K., Kirschner, P.A., Jochems, W., & Van Buuren, H. (2007). Measuring perceived sociability of computer-supported collaborative learning environments. *Computers & Education*, 49(2), 176-192.
- Kreijns, K., Kirschner, P.A., Jochems, W., & Buuren, H. v. (2004). Measuring perceived quality of social space in distributed learning groups. *Computers in Human Behavior*, 20(5), 607-632.
- Nijdam, B., & Van Buuren, H. (1979a). *Statistiek voor de sociale wetenschappen* (Vol. 1). Alphen aan den Rijn: Samsom Uitgeverij.
- Nijdam, B., & Van Buuren, H. (1979b). *Statistiek voor de sociale wetenschappen* (Vol. 2). Alphen aan den Rijn: Samsom Uitgeverij.
- Nijdam, B., & Van Buuren, H. (1988). *Statistiek voor de sociale wetenschappen. Beschrijvende statistiek* (Vol. 1). Alphen aan den Rijn: Samsom Uitgeverij.
- Nijdam, B., & Van Buuren, H. (1989). *Statistiek voor de economische wetenschappen* (Vol. 1). Alphen aan den Rijn: Samsom Uitgeverij.
- Nijdam, B., & Van Buuren, H. (1990). A Comparison of StatiOn with CAI and SPSS : Discussion of the Results. Paper presented at the Third International Conference on Teaching Statistics, Dunedin, New Zealand.
- Nijdam, B., & Van Buuren, H. (1991a). *Statistiek voor de economische wetenschappen* (Vol. 2). Alphen aan den Rijn: Samsom Uitgeverij.
- Nijdam, B., & Van Buuren, H. (1991b). *Statistiek voor de sociale wetenschappen. Inleiding in de inductieve statistiek* (Vol. 2). Alphen aan den Rijn: Samsom Uitgeverij.



- Nijdam, B., & Van Buuren, H. (1997). *Statistiek voor de sociale wetenschappen. Vergelijking groepen en regressie* (Vol. 3). Alphen aan den Rijn: Samsom Uitgeverij.
- Nijdam, B., & Van Buuren, H. (1999). *Statistiek voor de sociale wetenschappen 1. Beschrijvende statistiek*. Groningen: Wolters-Noordhoff.
- Nijdam, B., & Van Buuren, H. (1999). *Statistiek voor de sociale wetenschappen 2. Inleiding in de inductieve statistiek*. Groningen: Wolters-Noordhoff.
- Portier, S.J., & Van Buuren, J.A. (1993). A new interactive learning environment to study statistics: Theoretical base and evaluative results. In M. Valcke & F. Lockwood (Eds.), *New Developments in Educational Research in Relation to Distance Education Materials*. Heerlen: Open Universiteit Nederland.
- Portier, S. J., & Van Buuren, J. A. (1995). An interactive learning environment (ILE) to study statistics: Effects of prior knowledge on the use of embedded support devices. *European journal of Psychology and Education*, X(2), 197-207.
- Van Buuren, H. (1999). Competentiegericht onderwijs: wat betekent dat voor de student? *Swetenswaardigheden, Open Universiteit Nederland*, 8, 12-16.
- Van Buuren, H., & Giesbertz, W. (1999). Naar een virtueel onderzoekcentrum. Innovatie van het methoden en technieken van onderzoeksonderwijs. In K. Schlusmans, R. Slotman, C. Nagtegaal & G. Kinkhorst (Eds.), *Competentiegerichte leeromgevingen* (pp. 137-150). Utrecht: Lemma.
- Van Buuren, H., & Giesbertz, W. (1999). *Re-designing courses for electronic delivery*. Paper presented at the AECS, Vienna.
- Van Buuren, H., & Hummel, H. (1997). *Onderzoek de Basis*. Groningen: Wolters-Noordhoff.
- Van Buuren, J.A. (1986). Inleiding statistiek. In J.G. Bethlehem & W.J. Keller (Eds.), *Statistiek* (pp. 1-35). Utrecht: Stichting Teleac.
- Van Buuren, J.A. (2002). *Teaching Statistics and Research Methods in a Virtual Learning Environment*. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town (SA).
- Van Buuren, J.A. (2006a). *De onderzoekscompetentie bij Psychologie OUNL: enkele bevindingen*. Paper presented at the 33ste Onderwijs Research Dagen, Amsterdam, Nederland.
- Van Buuren, J.A. (2006b). *Teaching statistics and research methods: an integrated approach*. Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Bahia, Brazilië.
- Van Buuren, J.A., Baarda, D.B, De Goede, M.P.M., Schobre, J.H.L., & Giesbertz, W.J. (1994). *Methoden en technieken van onderzoek 1. Opzet en uitvoering*. Heerlen: Open Universiteit Nederland.
- Van Buuren, J.A., Bijker, M.M., & Wynants, G. (2005). *Instrumenten voor de effectmeting van competentiegericht onderwijs bij psychologiestudenten in afstandsonderwijs*. Paper presented at the 32ste Onderwijs Research Dagen, Gent, België.
- Van Buuren, J.A., & Giesbertz, W. J. (1998). *Innovatie M&T Onderwijs. Naar een Virtueel Onderzoek Centrum*. Heerlen: Open Universiteit Nederland.
- Van Buuren, J.A., Hutjes, J.M., & Giesbertz, W. (1990). *Kwalitatief onderzoek: de gevalsstudie als onderzoeksstrategie*. Heerlen: Open Universiteit Nederland.
- Van Buuren, J.A., Hutjes, J.M., Kropman, J.A., Loth, F., Dreteler, C., & Hoogveld, B. (1987). *Methoden en technieken van sociaal-wetenschappelijk onderzoek* (Vol. 2). Heerlen: Open Universiteit Nederland.
- Van Buuren, J.A., Hutjes, J.M., Kropman, J.A., Op de Weegh, P.U.M., Loth, F., Dreteler, C., et al. (1986). *Methoden en technieken van sociaal-wetenschappelijk onderzoek* (Vol. 1). Heerlen: Open Universiteit Nederland.

- Van Buuren, J.A., & Kreijns, K. (2003). *Het IMTO-project: Aanzet tot herinrichting van het psychologiecurriculum richting competentiegebaseerd onderwijs*. Paper presented at the 30ste Onderwijs Research Dagen, Kerkrade, Nederland.
- Van Buuren, J.A., Nijdam, A.D., Gastkemper, F.H.D., & Schlussmans, K.H.L.A. (1985a). *Statistiek 1a*. Heerlen: Open Universiteit Nederland.
- Van Buuren, J.A., Nijdam, A.D., Gastkemper, F.H.D., & Schlussmans, K.H.L.A. (1985b). *Statistiek 1b*. Heerlen: Open Universiteit Nederland.



# **Van vakgericht naar competentiegericht statistiekonderwijs**

*Een interventiestudie in een opleiding psychologie*





---

# Inhoud

## Voorwoord

- 1 Van problemen in het statistiekonderwijs naar het doel van de interventie**
  - 1.1 Inleiding
  - 1.2 Problemen in het statistiekonderwijs
    - 1.2.1 Problemen in het statistiekonderwijs: de student als verklaringsfactor
    - 1.2.2 Problemen in het statistiekonderwijs: het onderwijs als verklaringsfactor
    - 1.2.3 Problemen in het statistiekonderwijs: de plaats van het statistiekonderwijs in het psychologiecurriculum als verklaringsfactor
  - 1.3 Interventiedoel, vraagstelling en overzicht van het proefschrift
- 2 Van interventiedoel naar veranderdoelen**
  - 2.1 Inleiding
  - 2.2 Bepaling van de gewenste leeruitkomsten: gedragsdoelen
  - 2.3 Selectie van de determinanten
    - 2.3.1 Direct veranderbare externe determinanten
    - 2.3.2 Indirect veranderbare persoonlijke determinanten: procesvariabelen
  - 2.4 Matrix van veranderingsdoelen
- 3 Van veranderdoelen naar instructieaanwijzingen**
  - 3.1 Inleiding
  - 3.2 Theoretische methodieken, praktijken en instructieaanwijzingen
    - 3.2.1 Cognitive Load Theory
    - 3.2.2 Meaningful Learning
    - 3.2.3 Cognitive Apprenticeship Instruction Model (CAIM)
    - 3.2.4 Enkele empirische bevindingen van CAIM
  - 3.3 Theory of the Research Training Environment
  - 3.4 Matrix van instructieaanwijzingen
- 4 Van instructieaanwijzingen naar een geïntegreerd competentiegericht onderwijsontwerp**
  - 4.1 Inleiding
  - 4.2 Het onderwijsontwerp
    - 4.2.1 Four-Component Instructional Design Model (4C/ID)
    - 4.2.2 Redenen voor de keuze van het ontwerpmodel

- 4.3 Naar een blauwdruk voor het M&T- en statistiekonderwijs
  - 4.3.1 Onderzoekstaak als leertaak
  - 4.3.2 Omvang en inhoud van benodigde M&T- en statistiekkennis
  - 4.3.3 Onderwijsblauwdruk van het onderwijsprogramma voor de onderzoekscompetentie
  - 4.3.4 Inpassing van het ontwerp van de onderzoekscompetentie in het curriculum van de opleiding psychologie
- 4.4 Pilots met het competentiegericht onderwijsontwerp
  - 4.4.1 Inleiding
  - 4.4.2 Inhoud en doelstellingen van de pilots
  - 4.4.3 Methoden
  - 4.4.4 Resultaten
  - 4.4.5 Discussie en conclusie
- 5 Van ontwerp naar implementatie: crosssectionele bevindingen**
  - 5.1 Inleiding
    - 5.1.1 Onderzoeksvraag en hypothesen
  - 5.2. Methoden
    - 5.2.1 Respondenten
    - 5.2.2 Variabelen en hun operationalisering
    - 5.2.3 Procedure
    - 5.2.4 Analyseschema
  - 5.3 Resultaten
    - 5.3.1 Vergelijking van de twee leeromgevingen
    - 5.3.2 Toetsing van de mediatie in het model met SEM
  - 5.4 Conclusie en discussie
    - 5.4.1 Conclusie effecten van competentiegericht versus vakgericht statistiekonderwijs
    - 5.4.2 Beperkingen van het onderzoek
    - 5.4.3 Slotbeschouwing
- 6 Van ontwerp naar implementatie: longitudinale bevindingen**
  - 6.1 Inleiding
    - 6.1.1 Onderzoeksvraag
  - 6.2. Methoden
    - 6.2.1 Respondenten
    - 6.2.2 Meetinstrumenten
    - 6.2.3 Procedure
    - 6.2.4 Analyseschema
  - 6.3 Resultaten
  - 6.4 Conclusie en discussie
    - 6.4.1 Conclusie effecten van competentiegericht versus vakgericht statistiekonderwijs
    - 6.4.2 Beperkingen van het onderzoek
    - 6.4.3 Slotbeschouwing

---

## 7 Evaluatie: reflectie en discussie

- 7.1 Inleiding
- 7.2 Samenvatting en beantwoording van de onderzoeksvragen
- 7.3 Opbrengsten van het onderzoek
- 7.4 Beperkingen van het onderzoek
- 7.5 Aanbevelingen voor vervolgonderzoek
- 7.6 Aanbevelingen voor de praktijk van statistiekonderwijs nieuwe stijl
- 7.7 Tot slot

## Bijlage I.0 Stappen Intervention Mapping

Referenties

Samenvatting

Summary

Dankwoord

Lijst van eigen publicaties en presentaties in relatie tot het thema van dit proefschrift

*Cd-rom*

*Van vakgericht naar competentiegericht statistiekonderwijs*

## Bijlagen

### I Matrices

I.0	Stappen Intervention Mapping	7
I.1	Matrix van veranderdoelen	11
I.2	Matrix van instructieaanwijzingen	19
I.3	Matrix tien ontwikkelstappen van het 4C/ID-model met instructieaanwijzingen	21
I.4	Onderwijsblauwdruk onderzoekscompetentie bachelorfase OUNL (situatie 2007)	23

### II Meetinstrumenten Vragenlijst

II.1	Vragenlijst evaluatie IMTO-pilots	29
II.2	Vragenlijst evaluatie onderzoekscompetentie 2005	47
II.3	Vragenlijst evaluatie $T_1$ -meting	75

### III Voorbereidende analyses

III.1	Vervolgonderzoek naar nonrespons in het crosssectionele onderzoek	93
III.2	Voorstudie Raschanalyses van de oorspronkelijke Likertschalen	113
III.3	Kruisvalidatie van het determinantenmodel met Structural Equation Modeling	169
III.4	IJking $T_1$ - $T_2$ -items met Raschmodel	179





*Bijlage I.0*

## **Stappen Intervention Mapping**

Cel	A	B	C	D																				
1	Hoofdstuk	1 Probleem → Interventiedoel	2 Interventiedoel → Veranderdoelen	3 Veranderdoelen → Instructieaanwijzingen																				
2	Onderzoeksvraag	Wat zijn de problemen in het statistiekonderwijs en welke verklaringen worden daarvoor gegeven?	Welke determinanten bepalen het succes van het leren van statistiek?	Welke onderwijskundige instructieaanwijzingen zijn effectief om de gewenste veranderingen in de leeromgeving te bewerkstelligen die leiden tot succes van het leren van statistiek?																				
3	Intervention Mapping taken	Taak 1.1 Analyseer de problemen en mogelijkheden voor oplossingen. Taak 1.2 Formuleer gewenste program- mauitkomsten ('Interventie- doel').	Taak 2.1 Specificeer leeruitkomsten in gedragsdoelen Taak 2.2 Selecteer belangrijke en veranderbare determinanten van die gedragingen. Taak 2.3 Maak een matrix van veranderdoelen door gedragsdoelen te kruisen met determinanten van de gedragingen.	Taak 3.1 Zoek theorieën bij de veranderdoelen. Taak 3.2 Leid methodieken af uit de theorie en de onderzoeks- literatuur. Taak 3.3 Vertaal de methodieken in praktische technieken. Taak 3.4 Ga na of de technieken overeenstemmen met de veranderdoelen																				
4	Theorie	<table><tr><th>Problemen Statistiekonderwijs</th><th>Doel van interventie</th></tr><tr><td>- Studentkenmerken</td><td>- indirecte aanpak</td></tr><tr><td>- Onderwijskenmerken</td><td>- indirecte aanpak</td></tr><tr><td>- Leeromgevingskenmerken</td><td>- directe aanpak</td></tr></table>	Problemen Statistiekonderwijs	Doel van interventie	- Studentkenmerken	- indirecte aanpak	- Onderwijskenmerken	- indirecte aanpak	- Leeromgevingskenmerken	- directe aanpak	<table><tr><th colspan="2">Determinanten</th></tr><tr><td>Externe</td><td>Persoonlijke</td></tr><tr><td>Inhoud Didactiek Technol.</td><td>Attitude Motivatie Leerstrategieën</td></tr><tr><td>Gedragsdoelen</td><td>Veranderdoelen</td></tr></table>	Determinanten		Externe	Persoonlijke	Inhoud Didactiek Technol.	Attitude Motivatie Leerstrategieën	Gedragsdoelen	Veranderdoelen	<table><tr><th></th><th>Methodieken &amp; praktijken</th></tr><tr><td>Veranderdoelen</td><td>Instructieaanwijzingen</td></tr></table>		Methodieken & praktijken	Veranderdoelen	Instructieaanwijzingen
Problemen Statistiekonderwijs	Doel van interventie																							
- Studentkenmerken	- indirecte aanpak																							
- Onderwijskenmerken	- indirecte aanpak																							
- Leeromgevingskenmerken	- directe aanpak																							
Determinanten																								
Externe	Persoonlijke																							
Inhoud Didactiek Technol.	Attitude Motivatie Leerstrategieën																							
Gedragsdoelen	Veranderdoelen																							
	Methodieken & praktijken																							
Veranderdoelen	Instructieaanwijzingen																							
5	Empirisch	<p>Determinantenmodel (fig. 2.1)</p> <pre>graph LR     subgraph "Externe en persoonlijke determinanten"         direction TB         E1[Externe determinanten: inhoud, didactiek, techniek]         E2[Persoonlijke determinanten: leerstrategieën, leeromgeving, leerstijl]     end      subgraph "Determinanten"         direction TB         D1[Attitude: kennis, houding, gedrag]         D2[Motivatie: intrinsieke, extrinsieke, verwachting]         D3[Verwachting: zelfvertrouwen, zelfeffectiviteit, toewijding]     end      subgraph "Leerstrategieën"         direction TB         L1[Leerstijl]         L2[Leermethoden]         L3[Leertijd]         L4[Leeromgeving]         L5[Leerstijl]     end      subgraph "Leeruitkomsten"         direction TB         U1[Verstandelijke]         U2[Attitude]         U3[Motivatie]     end      E1 --&gt; D1     E1 --&gt; D2     E1 --&gt; D3     E2 --&gt; D1     E2 --&gt; D2     E2 --&gt; D3     D1 --&gt; L1     D1 --&gt; L2     D1 --&gt; L3     D1 --&gt; L4     D1 --&gt; L5     D2 --&gt; L1     D2 --&gt; L2     D2 --&gt; L3     D2 --&gt; L4     D2 --&gt; L5     D3 --&gt; L1     D3 --&gt; L2     D3 --&gt; L3     D3 --&gt; L4     D3 --&gt; L5     L1 --&gt; U1     L1 --&gt; U2     L1 --&gt; U3     L2 --&gt; U1     L2 --&gt; U2     L2 --&gt; U3     L3 --&gt; U1     L3 --&gt; U2     L3 --&gt; U3     L4 --&gt; U1     L4 --&gt; U2     L4 --&gt; U3     L5 --&gt; U1     L5 --&gt; U2     L5 --&gt; U3</pre>																						





## Matrix van veranderdoelen

Gedrags- doelen	Externe determinanten Leeromgeving			Persoonlijke determinanten Student: procesvariabelen			
	Inhoud	Didactiek	Techno- logie	Attitude	Motivatie	Leerstrate- gieën	Zelf- regulatie
1	I.1a-i	D.1a-i	T.1a-e	A.1a-e	M.1a-k	L.1a-e	Z.1a-e
2			T.2a-b		M.2a	L.2a-b	Z.2a
3	I.3a		T.3a				
4	I.4a	D.4a	T.4a	A.4a-b			Z.4a-h

Gedragdoelen (G) Student ...	Externe determinanten Leeromgeving	Inhoud (I)	Didactiek (D)	Technologie (T)
G.1 Toont beheersing van functionele kennis van de statistiek bij onderzoekso opdracht		<p>I.1.a. Verschuift accent op definitie-afleidingen en kansbepalingen naar doelen, voorwaarden voor gebruik en toepassingen van de statistiek in sociaalwetenschappelijk onderzoek.</p> <p>I.1.b. Behandelt statistische technieken vanuit vraag naar samenhang en naar verschil in de context van een psychologische probleemstelling en faciliteert daarmee de aansluiting van de cognitieve leerfase naar de associatieve en autonome leerfasen.</p> <p>I.1.c. Laat vanuit de theorie het verband zien van definiëren van concepten naar operationaliseren van variabelen en het daarmee samenhangend meetniveau, in relatie tot de daarmee verbonden statistische technieken.</p> <p>I.1.d. Laat consequenties zien van de keuze van het onderzoeksdesign voor de generalisatie van relatie(s) tussen variabelen en verbindt deze condities tevens met de concrete authentieke onderzoeksituatie.</p> <p>I.1.e. Laat consequenties zien van keuze voor (a)selecteren en randomiseren voor een experimenteel onderzoeksdesign / een verschilvraag of samenhangvraag in de authentieke leersituatie.</p>	<p>D.1.a. Positioneert en presenteert statistiek in betekenisvolle contexten, i.c. psychologisch onderzoek.</p> <p>D.1.b. Maakt een lange, ononderbroken leerlijn mogelijk en faciliteert daarmee praktisering en bevordering van retentie.</p> <p>"D.1.c. Faciliteert praktiseren en herhaling over een lange periode van onderzoekstaken gedurende de gehele bachelorfase."</p> <p>"D.1.d. Geeft student voldoende tijd, context en betekenis om statistisch netwerkschema op te bouwen. volgens drie achtereenvolgende fasen: 1. cognitief2. associatief en 3. autonoom."</p> <p>D.1.e. Behandelt statistiek niet in geïsoleerde positie, maar in haar relatie tot psychologische theorieën en onderzoeksmethoden.</p>	<p>T.1.a Activeert epistemische inter-acties door groepstaken (interdependentie, individuele verantwoordelijkheid, individuele doelen en/of rollen en groepsdoelen).</p> <p>T.1.b. Neemt door gebruik van statistische software het tijdrovende reken- en zoekwerk (bijvoorbeeld overschrijdingskansen bij toetsing) uit handen waardoor kans op 'learning space' vrij komt ten gunste van 'effectieve cognitieve belasting'.</p> <p>T.1.c. Stelt in de elektronische leeromgeving nieuwsgroepen en discussiegroepen in, ondersteund door vraagbaak en algemene nieuwsgroep, om het onderling discussiëren te faciliteren.</p> <p>T.1.d. Faciliteert samenwerking tussen studenten en docenten in de elektronische leeromgeving.</p> <p>T.1.e. Faciliteert docenten in het begeleiden van studenten in de elektronische leeromgeving.</p>

Gedragsdoelen (G) Student ...	Externe determinanten Leeromgeving		
	Inhoud (I)	Didactiek (D)	Technologie (T)
	<p>I.1f. Behandelt APA-regels hoe te rapporteren over resultaten van statistische analyses, waardoor het verband met de verslaggeving in de realistische onderzoekspraktijk wordt versterkt en in elk onderzoek weer opnieuw wordt gepraktiseerd.</p> <p>I.1g. Biedt een veelheid aan mogelijkheden tot interpreteren van analyseresultaten; maakt daarbij gebruik van authentieke cases met realistische data, waardoor relaties in het associatieve netwerk worden versterkt..</p>	<p>D.1f Legt in onderzoekscases verbinding met (toekomstige) psychologische praktijken.</p> <p>D.1g Faciliteert en stimuleert docenten psychologie, M&amp;T en statistiek tot samenwerking bij het opstellen van betekenisvolle onderzoekstaken en het begeleiden van studenten die onderzoekstaken vervullen en verbreedt de betekenis van statistiek tot statistisch denken en redeneren, tot het academische socialisatieproces van de student, wat vakoverstijgend is en consistent doorgezet moet worden in de psychologievakken.</p> <p>D.1h. Verbindt statistiek met psychologie en M&amp;T door statistiek te relateren aan voorafgaande vraagstellingen en methoden en de daaraan voorafgaande theorieën of theoretische conceptuele modellen.</p>	
	<p>I.1h. Behandelt statistische technieken voor diverse onderzoekssituaties, uitgaande van authentieke cases en faciliteert hiermee het praktiseren van de technieken en hun betekenis, waardoor de mogelijkheden tot aansluiting bij de associatieve en autonome leerfase worden vergroot.</p>		



Gedragsdoelen (G) Student ...	Externe determinanten Leeromgeving	Inhoud (I)	Didactiek (D)	Technologie (T)
		<p>I.1i. Vertrekt bij onderzoekstaken vanuit artikelen uit wetenschappelijke psychologietijdschriften en introduceert deze als uitdagend academisch onderzoeksleerdoel, eventueel in de vorm van een uitgewerkt voorbeeld. Het uitdagende leerdoel krijgt daardoor betekenis voor de student, wordt binnen het bereik van de taakopdracht gebracht en wordt daarmee haalbaar voor de student. Het specifieke academische onderzoeksdoel is dichterbij de ervaringswereld van de student gebracht en centraal gesteld als doel van onderzoekstaak</p> <p>I.1j. Behandelt condities waaronder statistische technieken mogen worden toegepast en contextualiseert deze, zodat de samenhang met de condities in de concrete onderzoekssituatie verhelderd wordt en betekenis krijgt</p>	<p>D.1i Brengt leerdoelen, instructie en assessment op één lijn en zorgt daarmee voor consistentie in opeenvolgende leerdoelen, leeracties en proeven van steeds toenemende bekwaamheid op het terrein van het doen van onderzoek</p>	

Gedragsdoelen (G) Student ...	Externe determinanten Leeromgeving		
	Inhoud (I)	Didactiek (D)	Technologie (T)
G.2 Beoordeelt vakliteratuur op statistische aspecten en communiceert hierover met vakgenoten			T.2a Laat studenten elkaars werkstukken, papiers commentariën (peer assessment en peer feedback). T.2b Waardeert peer reviews bij de beoordeling van uitgevoerde onderzoekopdrachten.
G.3 Toont inzicht in de samenhang tussen de wetenschappen	I.3a. Behandelt statistiek op een begrijpelijke, niet al te abstracte manier zonder wiskundige bewijsvoering en in haar relatie tot de psychologische onderzoeksvraag en daarbij passende methoden en technieken van onderzoek I.3b. Laat student statistisch arsenaal aan technieken opbouwen vanuit onderzoeksvragen in een variëteit aan onderzoekscases en onderzoekstaken		T.3a Biedt de groep de gelegenheid om als team de hele onderzoekstaak competent te vervullen, te onderhandelen over de taakverdelingen en het tijdsplan om een onderzoekstaak als teamproject invulling te geven. Verenigt onderzoekscompetenties met interpersoonlijke vaardigheden
G.4 Voert zelfstandig onderzoek uit (autonomie)	I.4a. Geeft ruimte voor eigen onderzoeksvragen	D.4a Maakt het mogelijk dat studenten op elk niveau – van simpel tot complex – onderzoek kunnen doen.	T.4a Maakt het studenten mogelijk om als 'community of inquiry' psychologisch onderzoek uit te voeren

Persoonlijke determinanten: procesvariabelen Student			
Attitude (A)	Motivatie (M)	Leerstrategieën (L)	Zelfregulatie (Z)
A.1a Is van mening dat statistiek ook leuk en interessant kan zijn.	M.1a Student zet zich in om zich statistiek eigen te maken en weet op de juiste momenten hulp te vragen, zodat hij/zij weer zelfstandig verder kan.	L.1a Beseft dat memoriseren een eerste fase in leerproces is en onvoldoende voor diepgaande verwerking van de leerstof.	Z.1a Plant eigen onderzoek door doelen te stellen en voorkennis te activeren.
A.1b Is van mening dat statistiek te beheersen is.	M.1b Voelt zich vertrouwd met toepassen van statistiek en toont geen tekenen van vermijding.	L.1b Brengt structuur aan in de statistische informatie.	Z.1b Stuurt de aandacht en richt deze op de onderzoekstaak.
A.1c Is van mening dat statistiek niet (te) moeilijk is.	M.1c Toont gedragsmatige betrokkenheid bij uitvoering van onderzoekstaak.	L.1c Probeert nieuwe kennis te incorporeren in de bestaande kennis door te parafraseren, samenvattingen te maken, analogieën, conceptuele modellen te gebruiken en aantekeningen te maken.	Z.1c Controleert zichzelf door zich vragen te stellen
A.1d Is van mening dat statistiek een waardevol onderdeel van het onderzoeks- proces is.	M.1d Toont affectieve betrokkenheid bij uitvoering van onderzoekstaak.	L.1d Probeert in elke nieuwe onderzoekssituatie aan te geven en te doorgronden wat de onderliggende statistische structuur is.	Z.1d Beheerst en reguleert de eigen inspanningen.
A.1e Is in zijn/haar selectieve perceptie positief gericht op de positie van statistiek in het onderzoeks- proces	M.1e Ontwikkelt interesse voor de statistische analyse van zijn onderzoeksdata en de samenhang van statistiek met psychologisch onderzoek. M.1f Beseft dat psychologisch onderzoek van belang is voor functioneren als 'scientist practitioner'. M.1f Beseft dat psychologisch onderzoek van belang is voor functioneren als 'scientist practitioner'.	L.1e Verifieert en toetst steeds de samenhangen tussen statistische output, de theorie en gebruikte M&T.	Z.1e Stelt zichzelf ten doel door te gaan met het volbrengen van de onderzoekstaak.

Persoonlijke determinanten: procesvariabelen Student			
Attitude (A)	Motivatie (M)	Leerstrategieën (L)	Zelfregulatie (Z)
<p>A.4a Is zeer betrokken bij het doen van onderzoek en daarom ook bij het toepassen van statistiek.</p> <p>A.4b Voelt zich competent om zelfstandig een gehele onderzoekstaak en de daarbij behorende statistische technieken uit te voeren.</p>	<p>M.1h Vindt onderzoekstaak belangrijk en nuttig.</p> <p>M.1i Beseft geen angst te hoeven hebben voor statistiek.</p> <p>M.1k Beseft de betekenis van de statistiek in het onderzoeks-proces.</p>	<p>L.2a Verwerkt onderzoeksinformatie systematisch en past dit adequaat toe bij alle voor de psychologische onderzoekspraktijk benodigde statistische technieken, zoals vereist voor de bachelorthesis.</p> <p>L.2b Corrigeert de eigen misconcepties, als verworven inzicht, voortvloeiend uit de epistemische uitwisselingen.</p>	<p>Z.2a Checkt de eigen cognities en corrigeert deze tijdens het onderzoeksproces.</p>
			<p>Z.4a Plant eigen onderzoek door doelen te stellen en voorkennis te activeren.</p> <p>Z.4b Stuur de aandacht en richt deze op de onderzoekstaak.</p> <p>Z.4c Controleert zichzelf door zich vragen te stellen</p> <p>Z.4d Beheerst en reguleert de eigen inspanningen.</p> <p>Z.4e Stelt zichzelf ten doel door te gaan met het volbrengen van de onderzoekstaak.</p>

Persoonlijke determinanten: procesvariabelen			
Student			
Attitude (A)	Motivatie (M)	Leerstrategieën (L)	Zelfregulatie (Z)
			Z.4f Checkt de eigen cognities en corrigeert deze tijdens het onderzoeksproces. Z.4g Monitort de voortgang en controleert de acties die nodig zijn om de juiste statistische bewerkingen uit te voeren en te interpreteren. Z4.h Weet tegenvallen te incasseren en door te zetten. Besteedt voldoende tijd en energie aan het uitvoeren van het onderzoek.

## Matrix van instructieaanwijzingen

<i>Globale instructieaanwijzing De onderwijsontwerper/docent</i>	<i>Nummer</i>
– Expliciteert de operationele doelstellingen, de vaardigheden die moeten worden aangeleerd.	1
– Zorgt dat de mentale belasting niet te zwaar wordt door de hoeveelheid nieuwe leerstof te beperken en voldoende aandacht te besteden aan de presentatie van het leermateriaal.	2
– Activeert de voorkennis van de lerende.	3
– Structureert de leerstof in op elkaar voortbouwende delen en gebruikt advance organizers.	4
– Laat op de instructie direct opdrachten volgen zodat kennis beter verwerkt en opgeslagen kan worden (responding en reproductie).	5
– Past veelvuldig steeds terugkerende vaardigheden toe: het steeds weer toepassen en herhalen van verworven kennis en vaardigheden heeft een groter leereffect dan wanneer leerstof er door veelvuldig memoriseren ingestampt moet worden.	6
– Biedt deeltaken aan voor routinematige, steeds terugkerende vaardigheden.	7
– Presenteert gevarieerde problemen en contexten om transfer van het geleerde te bevorderen.	8
– Kiest voorbeelden die consistent zijn met het leerdoel: voorbeelden en non-voorbeelden voor concepten, demonstraties voor procedures, visualisaties voor processen en modelling voor gedrag (Merrill, 2001).	9
– Richt het onderwijs zo in dat de docent én optreedt als inhoudelijk én als procesbegeleider, met andere woorden optreedt als coach, tutor, facilitator, 'zone van naaste ontwikkeling' en als les'gever'.	10
– Draagt er door het taakontwerp en de procesbegeleiding zorg voor dat de student zijn eigen leerproces binnen de grenzen van de (authentieke) casus zelf controleert.	11
– Selecteert leerdoelen vooral op het hogere niveau van oplossen van zogeheten slecht-gedefinieerde, complexe problemen.	12
– Stelt authentieke taken centraal die ontleend zijn aan de werkelijkheid i.c. uit de praktijk van praktiserende psychologen in het veld.	13
– Geeft een expertoplossing voor elke taak die nieuw is in zijn soort waarbij de lerenden de uitvoering van de taak observeren en zich aldus een beeld vormen van de processen die vereist zijn om de taak tot een goed einde te brengen (modelling).	14
– Brengt toenemende complexiteit aan in taken zodanig dat ze steeds meer vaardigheden en concepten bevatten die expertgedrag weerspiegelen.	15
– Vereenvoudigt of past het niveau van de taak aan of vermindert ondersteuning naarmate de lerende zelfstandiger wordt (scaffolding en fading).	16
– Zorgt voor voldoende variatie in taakuitvoering.	17
– Presenteert taken eerst in hun geheel vooraleer aandacht geschonken wordt aan de afzonderlijke onderdelen.	18
– Selecteert vooral activerende werkvormen die garanderen dat lerenden zowel experimenteren, observeren, reflecteren en abstraheren.	19
– Moedigt studenten aan hun kennis, manieren van redeneren en probleem oplossen te expliciteren (articulatie) naar docent en medestudent.	20
– Biedt tools en/of omgevingen aan om zelf actief informatie op te zoeken, te verwerken en te integreren.	21
– Stelt leerdoelen, instructie en toetsing op één lijn ('constructive alignment', Biggs, 2003) en controleert deze op adequate uitvoering.	22
– Ontwikkelt een leeromgeving die het verwerven van functionele (= declaratieve + procedurele + conditionele) kennis ondersteunt.	23
– Geeft als faculteit het voorbeeld voor wetenschappelijk gedrag en attitudes.	26
– Zorgt voor een positieve bekrachtiging van wetenschappelijke activiteit in de leeromgeving, zowel formeel als informeel.	27
– Betrekt de student vroeg in de opleiding en op een zo min mogelijk bedreigende manier.	28
– Benadrukt tijdens de training dat alle onderzoeksstudies hun beperkingen hebben.	29
– Laat studenten verschillende onderzoeksbenaderingen aanleren en naar waarde schatten.	30
– Toont studenten hoe wetenschap en praktijk kunnen samengaan.	31
– Stimuleert studenten om naar 'binnen' te kijken Laat studenten integraal deel uitmaken van het kennisgenererende proces.	32

<i>Globale instructieaanwijzing De onderwijsontwerper/docent</i>	<i>Nummer</i>
– Benadrukt wetenschap als een sociaal-interpersoonlijke ervaring.	33
– Stemt het statistiekonderwijs af op de behoeften en het niveau van wetenschappers ('scientist practitioners').	34
– Laat studenten praktijk opdoen.	35

## Tien ontwikkelstappen van het 4C/ID-model met instructieaanwijzingen

<i>Component 4C/ID</i>	<i>Stap</i>	<i>Ontwikkelstap</i>	<i>Instructieaanwijzingen*</i>
Leertaken	1	Ontleed de complexe vaardigheid (= competentie) in de samenstellende geïntegreerde vaardigheden en formuleer de criteria voor (acceptabele) uitvoering van de taak.	1, 10, 12, 30, 31, 34, 35
	2	Stel de volgorde van de taakklassen van eenvoudig naar complex vast.	3, 10, 15, 16
	3	Ontwerp de leertaak en zorg voor voldoende variatie tussen leertaken binnen een taakklasse.	2, 8, 9, 10, 13, 17, 18, 19, 22
Ondersteunende informatie (theorie, abstract)	4	Analyseer mentale modellen.	4, 10, 20, 23, 24, 26, 27, 29, 31, 32, 33
	5	Analyseer cognitieve strategieën.	10, 14, 20, 30
Just-in-time-informatie (to the point, context)	6	Ontwerp ondersteunende informatie.	10, 14, 21, 30, 34
	7	Analyseer regels en procedures.	10, 14, 21, 28, 30
	8	Analyseer vereiste kennis.	10, 11, 21, 25, 30, 31, 34
Deeltaakoefening	9	Ontwerp just-in-time-informatie.	6, 7, 10
	10	Ontwerp deeltaakoefeningen als extra ondersteuning voor het afzonderlijk trainen van complexe vaardigheden waarvoor een hoog niveau van automatisering gewenst of vereist is.	5, 6, 7, 10

\* (op basis van het doel van de interventie en de daarvan afgeleide gedragsdoelen)





## Onderwijsblauwdruk onderzoekscapaciteit Bachelorfase Psychologie OUNL (situatie 2007)

Competentie-niveau	Taakklasse Student	APA: Introductie	APA: Methoden	APA: Resultaten	APA: Discussie	Voorbeeld assessment
1A	Kan een eenvoudige secundaire analyse uitvoeren (op grote databestanden) waarbij groepen beschrijvend worden vergeleken op basis van frequentieverdeling, gemiddelde en standaardafwijking	Formuleren en verhelderen van onderzoeksvraag en theoretisch kader <ul style="list-style-type: none"> <li>- kan doelstelling en vraagstelling (over verschil tussen groepen) van onderzoek herformuleren in lijn met casus</li> <li>- kan vraagstelling uitwerken tot toetsbare hypothesen</li> </ul>	Operationaliseren en verzamelen van data <ul style="list-style-type: none"> <li>- kan in de gegeven casus aangeven welke variabelen onderzocht moeten worden om de vraagstelling te kunnen beantwoorden</li> <li>- kan de rol van de respondenten in de onderzoeken aangeven</li> </ul>	Analysen van data kan met <ul style="list-style-type: none"> <li>- frequentieverdeling</li> <li>- gemiddelde</li> <li>- standaardafwijking</li> <li>- een beschrijvende vergelijking van groepen maken.</li> </ul>	Evaluëren en rapporteren <ul style="list-style-type: none"> <li>- kan resultaten van onderzoek terugkoppelen naar vraagstelling</li> <li>- kan generalisatiemogelijkheden (interne en externe validiteit) van de bevindingen aangegeven</li> </ul>	<ul style="list-style-type: none"> <li>- kan een casus met groot databestand data theoretisch en methodisch (SPSS) voorbereiden en vervolgens op een mc-tentamen – waarvan de items voor een deel betrekking hebben op de casus en de daarop gebaseerde analyses – en een ander deel op de leerstof – met goed resultaat afsluiten (dit is een mc-casustoets)</li> </ul>
1B	Kan met behulp van t-toets een eenvoudige secundaire analyse uitvoeren (op grote databestanden) waarbij twee groepen met elkaar worden vergeleken	Als boven	Als boven	<ul style="list-style-type: none"> <li>- kan met t-toets voor twee groepen een inductieve / toetsende vergelijking tussen twee (gepaarde en ongepaarde) groepen maken</li> <li>- kan een betrouwbaarheidsinterval opstellen</li> </ul>	Als boven	Als boven
1C	Kan met behulp van eenvoudige variantieanalyse (ANOVA) een eenvoudige secundaire analyse uitvoeren (op grote databestanden) waarbij twee of meer groepen met elkaar worden vergeleken	Als boven	Als boven	<ul style="list-style-type: none"> <li>- kan met enkelvoudige variantieanalyse een inductieve/toetsende vergelijking tussen twee of meer (gepaarde en ongepaarde) groepen maken</li> <li>- kan post-hoc-analyses uitvoeren</li> <li>- kan aangeven wat de relatie is tussen t-toets en ANOVA</li> </ul>	Als boven	Als boven

Competentie-niveau	Taakklasse Student	APA: Introductie	APA: Methoden	APA: Resultaten	APA: Discussie	Voorbeeld assessment
1D	Kan met behulp van: bivariate regressieanalyse een eenvoudige secundaire analyse (op grote databestanden) uitvoeren	<ul style="list-style-type: none"> <li>- kan doelstelling en vraagstelling (over samenhang tussen 2 variabelen) van onderzoek herformuleren in lijn met casus</li> <li>- kan vraagstelling uitwerken tot toetsbare hypothese(n)</li> </ul>	<ul style="list-style-type: none"> <li>- kan in de gegeven casus aangeven welke variabelen onderzocht moeten worden om de vraagstelling te kunnen beantwoorden</li> <li>- kan een padmodel met 2 variabelen opstellen</li> <li>- kan de rol van de respondenten in de onderzoeken aangeven</li> </ul>	<ul style="list-style-type: none"> <li>- kan met correlatie de sterkte van een samenhang tussen twee variabelen bepalen</li> <li>- kan met lineaire regressie de richting van de samenhang tussen twee variabelen bepalen</li> <li>- kan op basis van analyseresultaten voorspellingen doen van de waarde op de afhankelijke variabele</li> </ul>	<p>Als boven</p> <p>Als boven</p>	<p>Als boven</p>
1A-D 5 EC				<ul style="list-style-type: none"> <li>- kan met correlatie de sterkte van een samenhang tussen twee variabelen bepalen en aangeven of er sprake is van interactie (moderatie of mediatie)</li> <li>- kan met lineaire regressie de richting van de samenhang tussen twee variabelen bepalen en aangeven of er sprake is van interactie (moderatie of mediatie)</li> <li>- kan aangeven hoe met ANOVA en correlatie/ regressieanalyse interactie kan worden aangetoond</li> </ul>	<p>als boven +</p> <ul style="list-style-type: none"> <li>- onvoorziene resultaten in een padmodel integreren</li> <li>- een kritische houding aannemen wat betreft eigen resultaten</li> <li>- kan eigen inzichten - kan generalisatie- mogelijkheden (interne en externe validiteit) van de bevindingen aangeven toepassen bij verklaren en interpreteren van resultaten</li> <li>- de bevindingen aangeven</li> </ul>	<ul style="list-style-type: none"> <li>- kan aan de hand van een uitgewerkt voorbeeld zelfstandig een compleet wetenschappelijk artikel volgens APA-regels in de Nederlandse taal schrijven</li> </ul>
2 5 EC	Kan een eenvoudig vragenlijstonderzoek (survey) opzetten en uitvoeren onder ten minste 20 respondenten	<ul style="list-style-type: none"> <li>- kan een onderzoeks-vraag formuleren waarin sprake is van (causale) samenhang tussen 3 of meer variabelen en waarin sprake is van interactie (resp. moderatie of mediatie)</li> <li>- kan vraagstelling uitwerken in een padmodel met minstens drie variabelen</li> </ul>	<ul style="list-style-type: none"> <li>- kan theoretische begrippen uit vraagstelling en/of hypothese(n) op valide wijze operationaliseren in meetinstrumenten die in een vragenlijst kunnen worden opgenomen</li> <li>- kan Likert-schalen construeren voor abstracte en complexe begrippen</li> <li>- kan de betrouwbaarheid (item-homogeniteit) van een Likertschaal item bepalen met Cronbach's ?</li> <li>- kan gedetailleerde planning maken om benodigde data te verzamelen</li> <li>- kan ten minste 20 respondenten werven, mondeling en/of via aanschrijfbrief</li> <li>- kan de rol van interviewer op zich nemen</li> </ul>	<ul style="list-style-type: none"> <li>- kan met correlatie de sterkte van een samenhang tussen twee variabelen bepalen en aangeven of er sprake is van interactie (moderatie of mediatie)</li> <li>- kan aangeven hoe met ANOVA en correlatie/ regressieanalyse interactie kan worden aangetoond</li> </ul>	<p>als boven +</p> <ul style="list-style-type: none"> <li>- onvoorziene resultaten in een padmodel integreren</li> <li>- een kritische houding aannemen wat betreft eigen resultaten</li> <li>- kan eigen inzichten - kan generalisatie- mogelijkheden (interne en externe validiteit) van de bevindingen aangeven toepassen bij verklaren en interpreteren van resultaten</li> <li>- de bevindingen aangeven</li> </ul>	

Competentie-niveau	Taakklasse Student	APA: Introductie	APA: Methoden	APA: Resultaten	APA: Discussie	Voorbeeld assessment
3 5 EC	Kan een literatuurstudie t.b.v. een onderzoeksplan uitvoeren Kan een eenvoudige secundaire analyse uitvoeren (op grote databestanden) waarbij groepen beschrijvend worden vergeleken op basis van nonparametrische data-analyse technieken	<ul style="list-style-type: none"> <li>- kan uit een probleem-schets een probleemstel-ling voor literatuuronder-zoek afleiden</li> <li>- kan zoekplan voor een methodische literatuur-studie opstellen</li> <li>- kan een methodische literatuurstudie uitvoeren</li> <li>- kan wetenschappelijke artikelen beoordelen op bruikbaarheid voor pro-bleemstelling</li> </ul>	<ul style="list-style-type: none"> <li>- kan in de gegeven casus aangeven welke varia-belen onderzocht moeten worden om de vraag-stelling te kunnen beantwoorden</li> </ul>	<ul style="list-style-type: none"> <li>- kan onderscheid aan-geven tussen de param-etrische en nonparam-etrische toetsen en deze adequaat kunnen in-zetten</li> <li>- kan de volgende toetsen toepassen: <ul style="list-style-type: none"> <li>• Chi-kwadraat toets</li> <li>• McNemar-toets</li> <li>• Mann-Whitney-toets</li> <li>• Kolmogorov-Smirnov-toets</li> <li>• Wilcoxon-toets</li> <li>• Kruskal-Wallis-toets</li> <li>• Logistische regressie</li> </ul> </li> <li>- alle tot nu toe behandelde statistische technieken +</li> <li>• MANOVA</li> <li>• herhaalde metingen</li> </ul>	<ul style="list-style-type: none"> <li>- als boven</li> </ul>	Voorbeeld assessment
4 5 EC	Een eenvoudig experi-ment opzetten en uit-voeren	<ul style="list-style-type: none"> <li>- kan doelstelling en vraagstelling (over verschil tussen groepen en samenhang tussen variabelen) van onder-zoek herformuleren in lijn met casus</li> <li>- kan vraagstelling uit-werken tot toetsbare hypothese(n)</li> </ul>	<ul style="list-style-type: none"> <li>- kan de rol van de res-ponden ten in de onder-zoeken aangeven</li> </ul>	<ul style="list-style-type: none"> <li>- kan resultaten van onderzoek terugkoppelen naar vraagstelling</li> <li>- kan generalisatie-mogelijkheden (interne en externe validiteit) van de bevindingen aangege-ven</li> </ul>	<ul style="list-style-type: none"> <li>- kan een compleet artikel schrijven volgens APA-regels Nederlandse taal, toevoegen van eigen data (n ? 12) aan bestaande set van experimentele data</li> </ul>	MC-casustoets
5 5 EC		<ul style="list-style-type: none"> <li>- kan experimentele opzetten beoordelen op interne en externe validiteit</li> <li>- kan theoretisch onder-bouwde experimentele vraagstelling formuleren en uitwerken tot experi-mentele hypothesen</li> </ul>	<ul style="list-style-type: none"> <li>- een gedetailleerde planning (experimen-teerschema) maken om data te verzamelen, rekening houdend met matching en randomi-satie</li> <li>- proefpersonen werven</li> <li>- als proefleider fungeren</li> </ul>			

Competentie-niveau	Taakklasse Student	APA: Introductie	APA: Methoden	APA: Resultaten	APA: Discussie	Voorbeeld assessment
6 5 EC	een kwalitatief onderzoek opzetten met gebruikmaking van diepte-interviews en observatie	<ul style="list-style-type: none"> <li>- kan aangeven wat de belangrijkste verschillen zijn tussen kwalitatief en kwantitatief onderzoek</li> <li>- kan aangeven welk soort vraagstellingen het beste beantwoord kunnen worden door middel van kwalitatief onderzoek</li> <li>- kan vanuit een probleemschets een probleemstelling formuleren en daaruit één of meer onderzoeks-vragen en deelvragen destilleren</li> </ul>	<ul style="list-style-type: none"> <li>- kan tot een verantwoordelijke keuze komen van de te onderzoeken eenheden</li> <li>- kan aangeven wat de verschillen zijn tussen de diverse soorten aselecte en niet aselecte steekproeven</li> <li>- kan bepalen hoe groot het aantal onderzoeks-eenheden dient te zijn</li> </ul>	<ul style="list-style-type: none"> <li>- kan aangeven hoe kwalitatieve data op een kwalitatieve manier worden verwerkt met als uiteindelijk doel de beantwoording van de probleemstelling</li> <li>- kan aangeven hoe output van onderzoek wordt omgezet in analyse-eenheden die passen bij het soort output</li> <li>- kan data prepareren voor analyse</li> <li>- kan een labelingsysteem opzetten en toepassen</li> <li>- kan aangeven hoe intersubjectiviteit gewaarborgd wordt</li> <li>- kan Cohen's kappa bepalen</li> </ul>	APA: Discussie	<p>Voorbeeld assessment</p> <ul style="list-style-type: none"> <li>- kan een casustentamen met open vragen met goed gevolg afsluiten</li> </ul>
7 13 EC	kan zelfstandig een empirisch-psychologisch onderzoek met eigen vraagstelling opzetten en uitvoeren volgens normen bachelorthesis	<ul style="list-style-type: none"> <li>- weet een compleet en theoretisch onderbouwd onderzoeksplan op te stellen</li> </ul>	<ul style="list-style-type: none"> <li>- kiest adequate M&amp;T en past deze correct toe</li> </ul>	<ul style="list-style-type: none"> <li>- kiest adequate data-analytische techniek(en) en past deze correct toe</li> </ul>		<ul style="list-style-type: none"> <li>- kan een onderzoek uitvoeren en erover rapporteren in de vorm van een Bachelor-thesis conform APA-regels, Nederlandse taal</li> <li>- kan de bevindingen rapporteren met een powerpointpresentatie</li> </ul>
Bachelor Psychologie Totaal 43 EC						

# Aanvulling onderwijsblauwdruk onderzoekscompetentie met niveauonafhankelijke vaardigheden

Niveau	Competentie Student	kennis en vaardigheden APA: Introductie	kennis en vaardigheden APA: Methoden	kennis en vaardigheden APA: Resultaten	kennis en vaardigheden APA: Discussie
algemeen	<input type="checkbox"/> is in staat wetenschappelijke problemen op te lossen <input type="checkbox"/> komt tot kritische distantie <input type="checkbox"/> heeft een academische houding	<input type="checkbox"/> kan probleem met een open benadering tegemoet treden <input type="checkbox"/> kan bestaande kennis toepassen bij definitie van een nieuw probleem <input type="checkbox"/> kan bekende principes gebruiken in nieuwe (probleem)situatie <input type="checkbox"/> kan beperkingen van theorie of model aangeven <input type="checkbox"/> kan essentie van probleem bepalen door hoofd- en bijzaken te onderscheiden <input type="checkbox"/> kan inspelen op (onverwachte) onderzoeksproblemen <input type="checkbox"/> kan een gedegen literatuurschets uitvoeren (ook Engelstalige literatuur)	<input type="checkbox"/> kan mbv verschillende media informatie en / of onderzoeksdata zoeken <input type="checkbox"/> begrijpt hoe metingen plaatsvinden <input type="checkbox"/> kan instructies en procedures volgen <input type="checkbox"/> kan op correcte wijze medewerking voor onderzoek verkrijgen <input type="checkbox"/> kan documentatie / voortgangsgeschiedenis bijhouden <input type="checkbox"/> kan keuzes met betrekking tot dataverzameling onderbouwen	<input type="checkbox"/> kan keuze voor analyse-techniek(en) onderbouwen <input type="checkbox"/> kan resultaten correct samenvatten <input type="checkbox"/> kan anticiperen op te verwachten resultaten <input type="checkbox"/> kan correcte conclusies trekken <input type="checkbox"/> kan onvoorziene resultaten in onderzoeksmodel integreren <input type="checkbox"/> kan kritische houding aan nemen t.o.v. resultaten van zichzelf en anderen <input type="checkbox"/> kan eigen inzichten toepassen bij verklaren en interpreteren van resultaten	<input type="checkbox"/> kan correct formuleren en spellen (correct taalgebruik) <input type="checkbox"/> kan onderzoek helder en gestructureerd rapporteren <input type="checkbox"/> kan presentatie aanpassen aan context en publiek <input type="checkbox"/> kan de essentie van het onderzoek schriftelijk en/of mondeling weergeven en verantwoorden <input type="checkbox"/> kan het onderzoek schriftelijk en/of mondeling presenteren <input type="checkbox"/> kan in discussie treden met collega's over onderzoek <input type="checkbox"/> kan vervolgonderzoek suggereren



## **Vragenlijst evaluatie IMTO-pilots**



**Vragenlijst ten behoeve van de evaluatie van de innovatie van het Methoden en Technieken onderwijs.**

Deze vragenlijst gaat in op diverse aspecten van het onderwijs zoals verzorgd in de pilots. De vragenlijst zal na elke pilot nagenoeg gelijk zijn. De vragen gaan in op zaken die in elke pilot aan de orde zijn: online leren, taakgericht onderwijs, (samenwerkend) leren, leerdoelen en communicatie.

De bedoeling van de vragen in deze vragenlijst is dat u aangeeft hoe u daar op dit moment, dus na het doorlopen van deze pilot, over denkt.

Het invullen van de vragenlijst zal naar verwachting 20 minuten duren.

De vragenlijst beoogt zowel meningen te inventariseren van studenten die de pilot hebben afgerond als van studenten die voortijdig zijn gestopt. Voor de laatste groep geldt dat de vragen kunnen worden ingevuld "voor zover van toepassing", d.w.z. voor zover u vindt dat u zich over betreffend onderdeel een oordeel heeft kunnen vormen.

Vult u eerst weer uw studentnummer in?

1. Studentnummer 83.....
2. Aan welke pilots heeft u deelgenomen, en welke heeft u afgerond?

	Deelgenomen		Afgerond	
	ja	nee	ja	nee
a Elektronische cursus statistiek				
b IMTO-basic				
c IMTO-I				
d IMTO-2				
e IMTO-3				
f Beta-run				
g Pilots binnen andere deelgebieden van de studie				
Psychologie				

**Voorbereiding en introductie op de pilot**

3. Waarom heeft u deelgenomen aan deze pilot?  
(1 = helemaal niet van toepassing, 5 = helemaal van toepassing)

	1	2	3	4	5
a Omdat ik samen met medestudenten kan studeren					
b Omdat ik kan leren omgaan met de computer, internet, discussiegroepen, netmeeting.					
c Om de integratie van inhoud met methode van onderzoek					
d Omdat ik aan nieuwe ontwikkelingen wil meedoen					
e Omdat ik goede ervaring heb met deelname aan eerdere pilots					
f Het is voor mij de enige manier om mijn certificaat te halen					
g Het meedoen aan pilots is voor mij financieel gezien interessant					
h Anders, namelijk					

4. Bent u van mening dat u gedurende de pilot voldoende geïnformeerd was in de volgende opzichten:

a Leerdoelen en doelstellingen	ja	nee
b Structuur en opzet van de pilot	ja	nee
c Samenwerkend leren	ja	nee
d Techniek (hardware en software)	ja	nee

## Communicatie

5. Onderstaande tabel geeft een overzicht van mogelijke vormen van communicatie waarvan u tijdens deze pilot wellicht gebruik heeft gemaakt. We vragen u van elk aan te geven of u er gebruik van heeft gemaakt - actief dan wel alleen passief (meelezen/meeluisteren) dan wel helemaal niet.

	<i>actief gebruikt</i>	<i>alleen passief gebruikt</i>	<i>niet gebruikt</i>
a Vraagbaak technische vragen algemeen			
b Vraagbaak inhoudelijke vragen algemeen			
c Om even met elkaar te babbelen algemeen			
d Om even met elkaar te babbelen algemeen			
e Vraagbaak GAW (groepen 1 en 2)			
f Chatprogramma (groepen 1 en 2)			
g E-mail communicatie met individuele medestudenten (via groepssite)			
h E-mail communicatie met individuele medestudenten (privé)			
i Telefonische communicatie met individuele medestudenten			
j E-mail communicatie met begeleiders (privé)			
k Telefonische communicatie met begeleiders (privé)			
l E-mail communicatie met de help-desk			
m Telefonische communicatie met de help-desk			
n Face-to-face contact met medestudenten (los van de introductiedag).			

- 
6. Bood het communicatiecentrum naar uw mening onvoldoende, voldoende of te veel interactie mogelijkheden *om de studietaak goed te doorlopen?*
1. onvoldoende
  2. voldoende
  3. te veel
7. Bent u van mening dat u voldoende feedback kreeg uit uw eigen subgroep?
1. ja
  2. nee
8. Bent u van mening dat u voldoende feedback kreeg van de begeleiding?
1. ja
  2. nee

### Online leren

9. Hoe ervaart u het studeren 'on-line' met de computer in vergelijking met het studeren op basis van gedrukt studiemateriaal?

	<i>meer</i>	<i>gelijk</i>	<i>minder</i>
a Belangstelling voor de inhoud			
b Begrip van de inhoud			
c Hulp van de studiebegeleider			
d Onzekerheid over de studieaanpak			
e Verlegenheid			
f Rust			
g Gemak van studeren			
h Zelfvertrouwen			
i Enthousiasme om te studeren			
j Zelf-evaluatie			
k Overzicht over de inhoud			

---

10. Vindt u onderstaande punten bezwaarlijk bij het werken aan de studie, zoals in deze pilot is gebeurd?



## Taakgericht onderwijs: taken en bronnen

12. In hoeverre bent u het eens met onderstaande stellingen? (1 = helemaal niet mee eens 5 = helemaal mee eens).

	1	2	3	4	5
a Werken met deeltaken geeft structuur aan het leerproces					
b Het beschikbaar zijn van bronnen* die geraadpleegd kunnen worden wanneer ik ze nodig heb, helpt de leerdoelen te bereiken					
c Studeren met een mediamix (schriftelijk materiaal, elektronisch materiaal, etc.) helpt de leerdoelen te bereiken					
d Werken met deeltaken helpt de leerdoelen te bereiken					
e Studeren met een mediamix resulteert in een verdieping van de leerstof					
f Het beschikbaar zijn van bronnen die geraadpleegd kunnen worden wanneer ik ze nodig heb, resulteert in een verdieping van de leerstof					
g Werken met deeltaken resulteert in een verdieping van de leerstof					
h Het beschikbaar zijn van bronnen die geraadpleegd kunnen worden wanneer ik ze nodig heb, is een stimulerende manier van leren					
i Werken met deeltaken is een stimulerende manier van leren					
j Studeren met een mediamix is een stimulerende manier van leren					
k Het beschikbaar zijn van bronnen die geraadpleegd kunnen worden wanneer dat nodig is, is een efficiënte manier van leren					
l Werken met deeltaken is een efficiënte manier van leren					
m Studeren met een mediamix is een efficiënte manier van leren					

\* Met bronnen wordt bedoeld: de reader, het SPSS-boek, boek: Statistiek in context, en de in de (algemene) bibliotheek geplaatste bronnen

### (Samenwerkend) leren

In onderstaande tabel leggen we u een aantal uitspraken voor over leren en samenwerkend leren. In hoeverre bent u het eens met onderstaande stellingen?  
(1 = helemaal niet mee eens / 5 = helemaal mee eens)

	1	2	3	4	5
a					
b					
c					
d					
e					
f					
g					
h					
i					
j					
k					
l					
m					
n					
o					
p					
q					
r					
s					

## Leerdoelen

14. In hoeverre bent u van mening dat deelname aan deze pilot voor u heeft bijgedragen aan de ontwikkeling van onderstaande kennis en vaardigheden (1. helemaal niet / 5. zeer sterk).

	1	2	3	4	5
a Zelfstandig leren					
b Samen met anderen aan taken werken					
c Kennis van statistische bewerkingen					
d Toepassen van statistische kennis					
e Inzicht in statistische probleemoplossingen					
f Kernachtig informatie verstrekken					
g Toepassen van statistische bewerkingen met SPSS					
h Zelfreflectie					
i Duidelijk en helder formuleren					
j Feedback geven/werk van medestudenten becommentariëren					
k Omgaan met commentaar en oordelen van medestudenten					
l Inzicht in eigen sterke en zwakke vaardigheden					
m Een eigen standpunt innemen en onderbouwen					
n Omgaan met Windows 95 of hoger					
o Omgaan met CD-rom					
p Omgaan met Tekstverwerking					
q Omgaan met E-mail					
r Omgaan met de mogelijkheden van de website discussieforums					
s Omgaan met Online leren					

In onderstaande tabel leggen we u een aantal uitspraken voor GAW. In hoeverre bent u het eens met onderstaande stellingen? (indien van toepassing) (1 = helemaal niet mee eens / 5 = helemaal mee eens)

	1	2	3	4	5
a Ik heb vanaf het begin niet geweten wat de bedoeling was van GAW					
b Ik heb vanaf het begin actief deelgenomen aan GAW					
c Mijn deelname aan GAW was vanaf het begin minder actief dan de anderen					
d Mijn deelname aan GAW was evenredig					
e Ik heb mij mbt GAW steeds passief opgesteld					
f Mijn deelname aan GAW werd steeds minder					
g GAW was voor mij een grote hulp.					
h Door GAW ontstond er een betere relatie onderling tussen de studenten					
i GAW irriteerde mij.					
j Er waren geen duidelijke afspraken te maken onderling mbt GAW					
k Door GAW konden wij onderling beter problemen oplossen					
l GAW had geen invloed op mijn studie.					
m Ik vind het vervelend dat GAW werd afgesloten					
n Door technische problemen werd ik geblokkeerd in het actief toepassen van de GAW					

---

**Statistiek II**

16. Bent u van mening dat u de leerdoelstellingen van het moduul Statistiek 2 hebt gerealiseerd?
  1. ja
  2. nee
  3. geen idee
17. Heeft u uw eigen doelen op het gebied van Statistiek en Methoden bereikt?
  1. ja
  2. nee
  3. geen idee
18. Bent u van mening dat u met deze wijze van studeren meer hebt geleerd van toepassingen van deelgebieden uit Statistiek II dan op een reguliere wijze?
  1. ja
  2. nee
  3. geen idee
19. Bent u van mening dat u nu, na deze IMTO 6, deelgebieden uit statistiek 1 en 2 zonder problemen in uw verdere studie kan toepassen?
  1. ja
  2. nee
  3. geen mening

Geef aan of u het eens of oneens bent met de volgende stellingen.

20. Het is me nu volkomen duidelijk hoe ik de samenhang tussen onafhankelijke en afhankelijke variabelen het beste kan toetsen en onderzoeken.
  1. eens
  2. oneens
  3. noch mee eens/noch mee oneens
21. Het is me nu volkomen duidelijk welke onderzoeksopzet het beste gehanteerd kan worden om de samenhang tussen variabelen te toetsen.
  1. eens
  2. oneens
  3. noch mee eens/noch mee oneens
22. Het is me nog steeds niet duidelijk wat regressie-analyse nu eigenlijk inhoudt.
  1. eens
  2. oneens
  3. noch mee eens/noch mee oneens



## Beoordeling

23. Welke aspecten van de pilot moeten volgens u meewegen in het eindoordeel?
1. Het cijfer wordt alleen bepaald door het eindproduct (de eindopdracht).
  2. Het cijfer wordt bepaald door het eindproduct en de opgeleverde tussen producten.
  3. Het cijfer wordt bepaald door het eindproduct en de bijdragen aan deelta-ken en de discussie.
  4. Het cijfer wordt bepaald door het eindproduct, de tussenproducten, de roluitvoering en de bijdragen aan de deeltaken en de discussie.

## Totaal indruk

24. Als u uw waardering voor deze pilot in een cijfer (van 1 t/m 10) zou moeten uitdrukken, welk cijfer zou u dan toekennen?
25. Zou u deze manier van studeren (on-line/samenwerkend leren) ook in uw verdere studietraject willen toepassen?
1. Ja
  2. Nee
26. Vindt u het voor uzelf wenselijk als onderstaande zaken bij de OU hoofdzakelijk beschikbaar zijn via internet?

Voorziening	Wenselijk	Onwenselijk
a Informatie over regelgeving		
b Informatie over cursussen en opleidingen		
c Oriënterende voorkennistoetsen		
l Oriëntatie op de OU-studie door het bestuderen van een proefpakket		
d Mogelijkheid tot inschrijven		
e Leermaterialen		
f In een team werken met medestudenten		
g Contact met begeleider-/mentor		
h Vaardigheden oefenen en feed-back krijgen		
i Eigen voortgang toetsen en feed-back krijgen		
j Naslagwerken en woordenboeken		
k Het tentamen uitwerken, doorsturen en afhandelen		

Geef uw mening over:

27. De bijgeleverde computerprogramma's (WinMax, Office, evt. GAW).  
Hoe was de aanlevering, het installeren, hulp bij problemen, het werken met de programma's enz.

This image shows a full page of white paper with horizontal dashed lines, typical of primary-ruled notebook paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

28. De Home Page van IMTO-6  
De "bereikbaarheid" (24 uur per dag?), het gebruik, overzicht ivm nieuwe berichten enz.

[illegible]

29. Het aanleveren van nieuwe opdrachten  
Duidelijkheid over de bedoeling van de opdrachten, doelstelling van de opdrachten enz.

[illegible]

### 30. Tijdschema

Duidelijkheid, het gebruik van het tijdschema, mogelijkheden om binnen het tijdsbestek de opdracht af te hebben, uiteindelijke duur van IMTO.

[illegible]

31. Begeleiding door de begeleiders.  
Bereikbaarheid, duidelijkheid, begrijpen van de vragen en antwoorden enz.

[illegible]

### 32. Algemene indruk van IMTO-6

[illegible]

27. Als u één ding mocht veranderen aan de opzet of uitwerking van deze pilot, wat zou u dan gewijzigd willen zien?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Hartelijk dank voor uw medewerking.





## **Vragenlijst evaluatie onderzoekscompetentie 2005**

**Vragenlijst Evaluatie Onderzoekscompetentie 2005**

In uw opleiding komt u cursussen tegen die betrekking hebben op het doen van onderzoek en de statistische verwerking van gegevens. Door middel van de vragen in deze vragenlijst willen we uw mening peilen over deze cursussen.

Daarnaast is het voor de ontwikkeling van het onderwijs en de afstemming daarvan op de behoeften van onze studenten belangrijk om te weten wat u op dit moment nog paraat hebt over de aangeboden leerstof. Deze vragen volgen na de vragen over leerstrategieën. Het zijn slechts negen vragen, waarvan drie meerkeuzevragen waarbij u uw antwoord kunt geven door een vakje aan te kruisen of te omcirkelen.

De overige zes parate kennisvragen zijn open vragen. Wij verzoeken u de vragenlijst zorgvuldig in te vullen.

Lees de vraag eerst rustig door en geef daarna antwoord.

Het invullen van de vragenlijst neemt wat betreft uw mening over het onderwijs ongeveer een half uur in beslag. De tijd voor het maken van de opgaven is zo beperkt mogelijk gehouden door het aanbieden van slechts negen opgaven en wordt mede bepaald door uw parate kennis van de leerstof over onderzoeksmethoden en statistiek en wat u daarover kwijt wilt.

U kunt de ingevulde lijst in een envelop zonder postzegels terugsturen naar het volgende adres:

Open Universiteit Nederland  
Directoraat Onderwijs – Psychologie  
t.a.v. drs. J.A. van Buuren  
Antwoordnummer 4400  
6400 VC Heerlen

Succes met uw studie!

Met vriendelijke groet,  
Hans van Buuren.

Om een administratie te kunnen houden, hebben wij van u drie gegevens nodig: uw studentnummer, uw postcode en uw huisnummer.

Vraag 001	Wat is uw studentnummer? Uw studentnummer bestaat uit 9 cijfers en begint met 83..	83.....
Vraag 002	Wat is uw postcode?	
Vraag 003	Wat is uw huisnummer?	

## Motivatatie en belangstelling voor onderzoek

De volgende items hebben betrekking op de wijze waarop u het *uitvoeren van onderzoek* tegemoet treedt.

Bij deze items vragen wij u naar uw motivatie en belangstelling voor het doen van onderzoek. Statistiek maar daar deel van uit, maar is niet het enige aspect. Onderzoek doen vindt plaats in een bepaalde context (bijvoorbeeld een psychologisch probleem of het toetsen van een psychologische theorie). Het omvat het uitvoeren van een literatuurstudie, het formuleren van de (psychologische) probleemstelling en onderzoeksvragen, het bepalen van de onderzoeksgroep en steekproefgrootte, de manier van data verzamelen, de selectie van de meetinstrumenten, enzovoorts. Kortom onderzoek doen omvat de vier fasen van de onderzoekscyclus: literatuuronderzoek, dataverzameling, data-analyse (statistiek) en rapportage.

Wij zijn dus geïnteresseerd in uw motivatie, betrokkenheid en belangstelling voor het ontwikkelen van praktische onderzoeksvaardigheden.

Voor deze items geldt dat er geen 'goede' of 'foute' antwoorden zijn. De antwoorden lopen uiteen van 1 = 'geheel oneens' tot 5 = 'geheel mee eens'. De items hebben betrekking op de wijze waarop u onderzoek uitvoert en daarbij uw kennis van statistiek en methoden toepast.

		geheel oneens	oneens	oneens noch eens	mee eens	geheel mee eens
004	Ik geef de voorkeur aan uitdagende onderzoekstaken zodat ik nieuwe dingen leer.	1	2	3	4	5
005	In deze modules put ik de meeste voldoening uit het begrijpen en doorgronden van de theorie en het tegelijkertijd toepassen van onderzoek.	1	2	3	4	5
006	Ik zou een onderzoekstaak kiezen die verbonden is met een interessante theorie waardoor ik er veel van kan leren, ook al betekent dit meer werk.	1	2	3	4	5
007	Ik probeer steeds van mijn fouten te leren, ook al krijg ik een goede beoordeling van mijn eindopdracht.	1	2	3	4	5
008	Ik denk dat er interessante dingen te leren zijn in deze modules.	1	2	3	4	5
009	Ik denk dat wat ik leer van statistiek door onderzoek te doen, ook kan gebruiken bij andere vakken.	1	2	3	4	5
010	Ik vind het belangrijk om zowel onderzoek te leren uitvoeren als statistiek toe te leren passen.	1	2	3	4	5
011	Ik ben zeer geïnteresseerd in het inhoudelijke aspect van onderzoek doen.	1	2	3	4	5
012	Ik denk dat wat ik allemaal door het uitvoeren van onderzoek leer, belangrijk is voor de rest van mijn studie.	1	2	3	4	5
013	Ik leer graag iets over onderzoeksmethoden en statistiek.	1	2	3	4	5
014	Begrip van statistische procedures is belangrijk voor mij.	1	2	3	4	5
015	Ik denk dat hoge cijfers voor de eindopdrachten van deze modules haalbaar zijn.	1	2	3	4	5
016	Ik ben er zeker van dat ik zelfs de moeilijkste onderzoekstaken en onderwerpen kan begrijpen.	1	2	3	4	5
017	Ik heb er vertrouwen in dat ik de basisstappen voor het uitvoeren van onderzoek kan begrijpen.	1	2	3	4	5
018	Ik ben er zeker van dat ik de meest complexe onderwerpen in deze onderzoeken kan begrijpen en daardoor kan vertalen in een gedegen onderzoeksaanpak.	1	2	3	4	5
019	Ik ben er zeker van dat ik de onderzoekstaken heel goed zal kunnen uitvoeren.	1	2	3	4	5
20	Ik verwacht dat ik deze modules met heel goede resultaten zal afronden.	1	2	3	4	5
21	Vergeleken met andere studenten die onderzoek doen en daarbij statistiek toepassen verwacht ik het beter te zullen doen.	1	2	3	4	5
22	Ik ben er zeker van dat ik deze modules goed zal kunnen invullen.	1	2	3	4	5

23	Ik denk dat ik een goede onderzoeker ben in vergelijking met andere studenten die onderzoek moeten doen.	1	2	3	4	5
24	Mijn onderzoeksvaardigheden zijn uitstekend vergeleken met die van andere studenten die onderzoek uitvoeren en daarbij statistiek toepassen.	1	2	3	4	5
25	Vergeleken met andere studenten in deze cursus(sen) denk ik dat ik veel weet over het uitvoeren van onderzoek.	1	2	3	4	5
26	Ik denk dat ik in staat ben om de theorie in deze modules te begrijpen en daardoor onderzoek uit te kunnen voeren.	1	2	3	4	5
27	Als ik mijn eindopdracht maak denk ik eraan hoe slecht ik het doe.	1	2	3	4	5
28	Ik heb een ongemakkelijk en ellendig gevoel als ik een eindopdracht moet maken.	1	2	3	4	5
29	Ik ben zo gespannen tijdens het maken van mijn eindopdracht, dat ik de dingen die ik geleerd heb niet meer op de juiste wijze toe kan passen.	1	2	3	4	5
030	Ik maak me zorgen over de eindopdrachten.	1	2	3	4	5
031	Als ik me voorbereid op een eindopdracht, dan zeg ik de feiten uit de theorie voor mezelf steeds hardop.	1	2	3	4	5
032	Wanneer ik het theoretisch gedeelte uit dit moduul lees dan herhaal ik de kernpunten keer op keer hardop voor mezelf zodat ik ze onthoud.	1	2	3	4	5
033	Om mezelf te helpen de theoretische informatie te onthouden lees ik mijn aantekeningen telkens over.	1	2	3	4	5
034	Als ik de theorie en onderzoekstaken bestudeer probeer ik dat met informatie uit andere bronnen te integreren.	1	2	3	4	5
035	Tijdens het bestuderen van de theorie probeer ik de dingen die ik lees, te verbinden met wat ik al weet.	1	2	3	4	5
036	Tijdens het bestuderen van het theoretisch gedeelte van de taken probeer ik belangrijke kwesties in mijn eigen woorden te formuleren.	1	2	3	4	5
037	Als ik een onderzoekstaak bestudeer, probeer ik dat met informatie uit verschillende bronnen te verbinden.	1	2	3	4	5
038	Bij het uitvoeren van nieuwe onderzoekstaken maak ik gebruik van wat ik heb geleerd in andere opdrachten.	1	2	3	4	5
039	Ik probeer me dingen die in het theoretisch gedeelte van de onderzoekstaken voorkomen voor te stellen.	1	2	3	4	5
040	Ik bedenk concrete voorbeelden bij de theorie.	1	2	3	4	5
041	Wat ik in de theorie leer betrek op mijn eigen privé en onderzoekservaringen.	1	2	3	4	5
042	Ik denk na of de theorie ook van toepassing is op mijn dagelijkse leven.	1	2	3	4	5

		geheel oneens	oneens	oneens noch eens	mee eens	geheel mee eens
043	Als hulp bij het bestuderen van het theoretisch gedeelte maak ik een samenvatting van de kernpunten daaruit.	1	2	3	4	5
044	Ik vind het moeilijk om de hoofdpunten te ontdekken in de theorie.	1	2	3	4	5
045	Ik maak tabellen, diagrammen of overzichten om de theorie en mijn onderzoeksplan beter te structureren.	1	2	3	4	5
046	Ik probeer de theoretische informatie zo te ordenen dat ik het goed in me op kan nemen.	1	2	3	4	5
047	Bij het bestuderen van het theoretische gedeelte probeer ik me te herinneren wat op de bijeenkomsten is gezegd, zodat ik de betekenis en achtergronden van het onderzoek zo goed mogelijk begrijp.	1	2	3	4	5
048	Tijdens het bestuderen van de theorie probeer ik altijd te begrijpen wat er precies bedoeld wordt, zelfs al is het niet direct van belang.	1	2	3	4	5
049	Als ik een eindopdracht voorbereid probeer ik mezelf zo veel mogelijk dingen uit het theoretisch gedeelte te herinneren.	1	2	3	4	5
050	Ik vraag me nooit af of de methode die ik hanteer bij het uitvoeren van mijn onderzoek de juiste is voor het beantwoorden van mijn onderzoeksvragen.	1	2	3	4	5
051	Altijd als theoretische concepten in empirisch onderzoek worden getoetst, vraag ik me af of manier van dataverzameling de onderzoeksresultaten vertekent.	1	2	3	4	5
052	Ik zie de onderzoeksopdrachten als startpunt en probeer er mijn eigen ideeën over te ontwikkelen.	1	2	3	4	5
053	Ik probeer te spelen met mijn eigen ideeën over onderzoek, gebaseerd op mijn ervaringen ermee.	1	2	3	4	5
054	Ik denk nooit na over mogelijke alternatieven voor beweringen of conclusies in onderzoeksverslagen die ik lees.	1	2	3	4	5
055	Het is voor mij heel lastig om tegenstrijdige conclusies uit verschillende onderzoeken te hanteren.	1	2	3	4	5
056	Ik focus altijd op de kwaliteit van de meetinstrumenten die men in een onderzoek gebruikt en ga na hoe betrouwbaar en valide die meetinstrumenten zijn.	1	2	3	4	5
057	Bij het lezen van onderzoeksverslagen kijk ik altijd meteen naar de grootte van de onderzoeksgroep om te zien of de conclusies daardoor wel voldoende ondersteund worden.	1	2	3	4	5
058	Bij psychologische concepten die empirisch worden onderzocht vraag ik me altijd af met welke theorieën ze verbonden zijn.	1	2	3	4	5

058	Bij psychologische concepten die empirisch worden onderzocht vraag ik me altijd af met welke theorieën ze verbonden zijn.	1	2	3	4	5
059	Als ik onderzoeksverslagen lees kijk ik nooit naar de beschrijving van de onderzoeksgroep.	1	2	3	4	5
060	Onderzoeksverslagen lees ik altijd kritisch.	1	2	3	4	5
061	Ik denk altijd na over mogelijke alternatieven voor uitspraken of conclusies in mijn eigen onderzoeksverslagen.	1	2	3	4	5
062	Ik stel mezelf vragen om er zeker van te zijn dat ik de theorie die ik heb bestudeerd, beheers.	1	2	3	4	5
063	Als een onderzoekstaak niet verplicht is, dan maak ik hem toch.	1	2	3	4	5
064	Voordat ik met de onderzoekstaak begin, denk ik na over de theorieën die ermee verbonden zijn zodat ik de onderzoeksstappen kan bepalen.	1	2	3	4	5
065	Als ik bij het uitvoeren van onderzoek niet alles begrijp dan maak ik even pas op de plaats om de theorie eromheen nog eens door te nemen.	1	2	3	4	5
066	Ik constateer vaak dat ik de theoretische en praktische informatie in deze cursus heb doorgelezen maar niet begrijp waar het eigenlijk over gaat.	1	2	3	4	5
067	Ik kom er vaak achter dat als ik bezig ben met het theoretisch gedeelte, dat ik met mijn gedachten ergens anders ben en daardoor eigenlijk niet echt de stof in me opneem.	1	2	3	4	5
068	Ik word vaak zo moe of verveeld als ik met deze cursus bezig ben dat ik er mee stop voordat ik afgerond heb wat ik van plan was.	1	2	3	4	5
069	Ik werk hard om een hoog cijfer te halen, zelfs al vind ik een onderzoeksopdracht en de theorie niet leuk.	1	2	3	4	5
070	Als de onderzoekstaak en de bijbehorende theorie moeilijk zijn geef ik het óf op óf ik doe alleen de eenvoudige delen.	1	2	3	4	5
071	Ik werk door totdat ik klaar ben, zelfs al zijn de taken saai en vervelend.	1	2	3	4	5
072	Als ik deelneem aan een bepaalde module dan zorg ik er ook voor dat ik die afmaak.	1	2	3	4	5
073	Als het nodig is leer ik ook 's avonds laat en in het weekeind.	1	2	3	4	5
074	Gewoonlijk duurt het niet erg lang voordat ik met het leren en de taken ga beginnen.	1	2	3	4	5
075	Ik prent de theorie en de statistische begrippen in mijn hoofd door steeds maar weer te herhalen.	1	2	3	4	5
076	Ik probeer van mijn fouten te leren, zelfs al presteer ik slecht bij een eindopdracht.	1	2	3	4	5



		geheel oneens	oneens	oneens noch eens	mee eens	geheel mee eens
077	Als ik onderzoek doe, ga ik helemaal op in het onderzoeksproces.	1	2	3	4	5
078	Het fascineert me om onderzoek te doen naar psychologische concepten.	1	2	3	4	5
079	Het aan den lijve ondervinden hoe psychologische concepten empirisch onderzocht worden, vind ik een van de boeiendste processen om mee te maken.	1	2	3	4	5
080	Statistiek boeit me niet, maar onderzoek doen vind ik wèl leuk.	1	2	3	4	5
081	Onderzoek doen boeit me niet, maar statistiek vind ik een leuk vak.	1	2	3	4	5

## Leerstrategieën

Hier volgen enkele uitspraken over uw onderzoekshouding en hoe u het uitvoeren van onderzoek combineert met kennis van psychologische theorieën en concepten. Er is geen *juiste* manier van studeren en onderzoek doen. Hoe er gestudeerd wordt en onderzoek wordt uitgevoerd hangt af van uw eigen stijl en het onderwijs dat wordt gevolgd. Net als de vorige items hebben deze stellingen betrekking op het doen van onderzoek. In welke mate zijn de uitspraken voor u waar?

U kunt hier kiezen van 1 = 'nooit of bijna nooit waar' tot 5 = 'altijd of bijna altijd waar'.

		nooit of bijna nooit waar	soms waar	de helft van de tijd waar	vaak waar	altijd of bijna altijd waar
082	Ik vind dat het ontwikkelen van vaardigheden in combinatie met verdieping in de theorie mij een gevoel van diepe persoonlijke voldoening geeft.	1	2	3	4	5
083	Ik vind dat ik me genoeg moet verdiepen in een onderwerp om mijn eigen conclusies te kunnen trekken: eerder ben ik niet tevreden.	1	2	3	4	5
084	Ik wil voor deze modules slagen met een minimum aan inspanning.	1	2	3	4	5
085	Eigenlijk bestudeer ik alleen de theorie en de praktische informatie die ik als leermateriaal bij deze module aangereikt krijg.	1	2	3	4	5
086	Ik heb het gevoel dat haast ieder onderwerp heel interessant kan zijn, als ik me er maar in verdiep.	1	2	3	4	5

087	Ik vind de meeste nieuwe onderwerpen interessant en dus besteed ik er extra tijd aan om er meer informatie over te vinden.	1	2	3	4	5
088	Ik vind deze cursus niet erg interessant, dus beperk ik mijn werk tot het minimum.	1	2	3	4	5
089	Ik stamp sommige dingen erin door ze te lezen en te herlezen tot ik ze uit mijn hoofd ken ook al begrijp ik ze niet.	1	2	3	4	5
090	Ik vind dat een onderzoek doen, als onderdeel van een academische studie, even boeiend kan zijn als een goede roman of film.	1	2	3	4	5
091	Ik test mezelf op belangrijke onderwerpen tot ik ze volledig begrijp.	1	2	3	4	5
092	Ik vind dat ik gemakkelijker door de eindopdrachten heen kom wanneer ik de belangrijkste onderdelen weet op te zoeken dan wanneer ik probeer de stof te begrijpen.	1	2	3	4	5
093	Gewoonlijk beperk ik dit onderdeel van mijn studie tot wat uitdrukkelijk aangegeven is omdat ik denk dat het niet nodig is om iets extra's te doen.	1	2	3	4	5
094	Ik werk hard aan dit onderdeel van mijn studie omdat ik de stof en de praktische toepassing ervan interessant vind.	1	2	3	4	5
095	Ik besteed een groot deel van mijn vrije tijd aan het zoeken naar extra informatie over de onderwerpen die in de onderzoeksmodule behandeld worden.	1	2	3	4	5
096	Ik vind het nutteloos onderzoeksonderwerpen diepgaand te bestuderen. Het is verwarrend (en tijdrovend) terwijl je toch alleen maar een oppervlakkige kennis van het onderwerp nodig hebt om je onderzoeksvaardigheden te ontwikkelen.	1	2	3	4	5
097	Ik vind dat docenten niet mogen verwachten dat studenten enorm veel tijd besteden aan het bestuderen van de theorie waarvan bekend is dat die toch een ondergeschikte rol speelt in de eindopdracht.	1	2	3	4	5
098	De meeste lessen over onderzoek doen roepen bij mij vragen op, waar ik een antwoord op wil.	1	2	3	4	5
099	Ik vind het belangrijk om de meeste literatuur die tijdens lessen of in de modules aanbevolen wordt, te raadplegen.	1	2	3	4	5
100	Ik vind het zinloos om zaken te bestuderen die waarschijnlijk geen rol spelen bij de eindopdracht.	1	2	3	4	5
101	Ik vind dat de beste manier om een goede beoordeling te krijgen voor je eindopdracht erin bestaat dat je goed weet waar je op het Net moet zoeken om informatie te kopiëren voor je verslag.	1	2	3	4	5

## Leren als concept

### Vraag 102

Wat verstaat u eigenlijk onder 'leren'?

Mensen verschillen niet alleen in leerstrategieën, maar ook in de opvattingen die zij hebben over leren: de leerconceptie. In de onderzoeksliteratuur ziet men een zestal verschillende opvattingen over leren. Deze staan hieronder vermeld.

Wilt u deze concepten rangordenen naar de mate waarin ze aansluiten bij uw persoonlijke rangordening (van 1 als hoogste tot 6 als laagste)?

In het invulvak onder de rangordeningen kunt u uw keuze toelichten.

*Rangorde*

102a Leren als het opnemen van informatie: ☐

103a Leren als memoriseren, het van buiten leren: ☐

104a Leren als het selecteren van informatie die nu of later ` kan worden toegepast: ☐

105a Leren als het achterhalen van de boodschap van een auteur; dingen anders zien: ☐

106a Leren als het ontwikkelen van competenties; het toepassen van kennis in nieuwe situaties. ☐

107a Leren als zelfontplooiing: ☐

Toelichting:

.....

.....

.....

.....

.....

.....

.....

.....

## Evaluatie van de eigen ervaringen in de onderzoeksmodules

### Vraag 108

Wijkt de manier waarop u zich voorbereidt op een eindopdracht voor statistiek of een werkstuk als afronding van een onderzoek af van de manier waarop u andere cursussen van de OUNL volgt?

.....

.....

.....

.....

### Vraag 109

Wat denkt u: weerspiegelt een behaald cijfer voor de onderzoeksmodules dat wat u hebt geleerd?

.....

.....

.....

.....

### Vraag 110

Sluit het onderwijs in de onderzoeksmodules goed aan bij de eindopdracht, of had u een andere toets verwacht?

.....

.....

.....

.....

**Vraag 111**

Voor een onderzoeksmodule staat 120 uur. Kunt u aan de hand van de hiernaast weergegeven opties aangeven of dat voldoende, te veel of juist te weinig is om zo'n module te voltooien?

- ☐ aanzienlijk minder dan 120 uur
- ☐ Minder dan 120 uur
- ☐ 120 uur
- ☐ Meer dan 120 uur
- ☐ Aanzienlijk meer dan 120 uur

Extra opmerkingen bij studieduur:

.....

.....

.....

.....

**Vraag 112**

Cursus statistiek 1 en/of 2 voltooid?

- ☐ Cursus(sen) statistiek succesvol afgerond.
- ☐ Cursus(sen) statistiek bestudeerd, maar nog geen tentamen gedaan.
- ☐ Cursus(sen) statistiek gedeeltelijk bestudeerd.
- ☐ Cursus(sen) statistiek nog niet ingekeken.
- ☐ Vrijstelling.

**Individuele karakteristieken in context**

Bepaalde elementen in uw onderzoeksbenadering zijn karakteristiek voor u als student in deze onderzoeksmodules. Het gaat bij de volgende vragen en stellingen dus om gedrag, dat u als voorkeursstijl hanteert bij het uitvoeren van onderzoek en het bestuderen van de bijbehorende theorie. Er is dus geen sprake van een "goed" of een "fout" antwoord, u geeft louter en alleen weer welke uitspraken kenmerkend zijn voor uw gedrag in deze onderwijscontext. U hebt daarbij de keuzemogelijkheid uit vijf antwoorden per item, variërend van 'helemaal niet van toepassing' tot 'helemaal van toepassing'. Kies het antwoord dat het meest overeenstemt met uw eigen manier van werken en denk niet te lang na over uw antwoord. Sla svp geen stellingen over.

		hele- maal niet van toepas- sing	weinig van toepas- sing	matig van toepas- sing	grote- deels van toepas- sing	hele- maal van toepas- sing
113	In de onderzoeksmodules neem ik altijd zelf de verantwoording voor mijn eigen acties en de consequenties daarvan..	1	2	3	4	5
114	Bij onderzoek doen neem ik steeds eigen initiatieven.	1	2	3	4	5
115	Op het gebied van empirisch onderzoek ben ik steeds erg goedgegelovig, omdat ik ervan uitga dat de reputatie van onderzoekers garant staat voor de betrouwbaarheid en geloofwaardigheid van hun resultaten.	1	2	3	4	5
116	Ik durf altijd mijn eigen vorm te geven aan een onderzoeksopdracht, ook al betekent dit dat ik het risico loop op een lager cijfer.	1	2	3	4	5
117	Ik bekijk een probleem en de oplossing daarvan altijd van verschillende kanten	1	2	3	4	5
118	Bij onderzoek volg ik altijd precies de voorgeschreven richtlijnen en methoden uit het werkboek.	1	2	3	4	5
119	Tijdens het uitvoeren van onderzoek neem ik altijd snel beslissingen, en waak ervoor dat dit niet ten koste gaat van de wetenschappelijke criteria voor onderzoek doen.	1	2	3	4	5
120	Ook al loop ik tegen onvoorziene problemen op in mijn onderzoek, dan beden ik daar toch een goede oplossing voor.	1	2	3	4	5
121	Bij onderzoek doen houd ik me altijd letterlijk aan de opdracht zoals die in het werkboek is omschreven, dan ben ik er zeker van dat ik het doe zoals het hoort.	1	2	3	4	5
122	Hoe meer wetenschappers een bepaald idee aanhangen, hoe makkelijker ik me aansluit bij hun ideeën en theorieën.	1	2	3	4	5
123	Ik tuin er altijd in bij misleidende grafiekjes.	1	2	3	4	5
124	Als ik eenmaal mijn plan heb gemaakt voor mijn onderzoek dan houdt niets of niemand mij tegen in de uitvoering ervan.	1	2	3	4	5
125	Door informatie te kopiëren uit officiële onderzoeksartikelen probeer ik altijd in mijn eigen eindopdracht letterlijk weer te geven wat wetenschappers al over mijn onderwerp hebben gezegd.	1	2	3	4	5
126	Bij onderzoek doen heb ik steeds ondersteuning nodig bij het nemen van beslissingen, anders kan ik niet verder.	1	2	3	4	5
127	Als ik met onderzoek bezig ben, borrelen er allerlei ideeën in mij op, om te verwerken in dat onderzoek.	1	2	3	4	5
128	In een discussie over onderzoek doen laat ik me snel intimideren door de argumenten van anderen.	1	2	3	4	5

129	Als ik een nieuw idee van mezelf wil toetsen in een onderzoek, dan doe ik dat altijd pas als ik een voorbeeld kan vinden in een ander onderzoek.	1	2	3	4	5
130	Ik vind onderzoek doen zo complex, dat ik nooit van de voorgeschreven paden afwijk, uit angst dat ik het dan niet meer overzie.	1	2	3	4	5
131	Ik houd ervan om in een discussie mijn eigen ideeën over onderzoek doen te toetsen aan die van een ander.	1	2	3	4	5

### Waarnemen en aandacht in de onderzoekscontext

De mate van aandacht die we besteden aan zaken die we al of niet de moeite waard vinden varieert naar gelang we vermoeider zijn, het moment van de dag, we meer interesse hebben in een taak enz. De volgende stellingen houden rekening met die variatie, maar u kunt ermee aangeven wat doorgaans het meest karakteristiek is voor u als zich bezig houdt met onderzoek. Ook hier is er geen sprake van “goede” of “foute” antwoorden, het gaat louter om de manier waarop u doorgaans uw aandacht verdeelt en richt op onderzoekskwesties. U hebt daarbij de keuzemogelijkheid uit vijf antwoorden per item, variërend van “helemaal niet van toepassing” tot “helemaal van toepassing”. Kies het antwoord dat het meest overeenstemt met de manier waarop u uw aandacht verdeelt en denk niet te lang na over uw antwoord. Sla svp geen stellingen over.

		helemaal niet van toepassing	weinig van toepassing	matig van toepassing	grooten- deels van toepassing	helemaal van toepassing
132	Ik vind het gemakkelijk om mijn volledige concentratie gericht te houden op het onderzoek waar ik op dat moment mee bezig ben.	1	2	3	4	5
133	Ik neem aandachtig mijn onderzoeksactiviteiten in mijn omgeving waar en ruis uit mijn omgeving verstoort dat niet.	1	2	3	4	5
134	Ik ben helemaal in het hier en het nu van mijn onderzoek en laat mijn gedachten niet afdwalen naar zaken uit het verleden of mogelijke toekomstige situaties.	1	2	3	4	5
135	In onderzoek doe ik dingen vaak automatisch, zonder dat ik er echt aandacht aan besteed.	1	2	3	4	5
136	Ik ben met mijn volledige aandacht bij mijn onderzoekstaak, en me zeer bewust van wat ik aan het doen ben.	1	2	3	4	5

---

## Kennis van onderzoeksmethoden en statistiek

Op de volgende bladzijden komen een aantal vragen over allerlei aspecten die te maken hebben met het doen van onderzoek en de statistische verwerking van gegevens. We willen door middel van deze vragen te weten komen wat u op dit moment weet over onderzoek en statistische analyse.

Lees de vraag eerst rustig door en geef daarna antwoord. Bij een drietal vragen kunt u een van de antwoordalternatieven kiezen door uw keuze te omcirkelen of aan te kruisen. Bij twee van deze keuzevragen krijgt u de mogelijkheid om een eigen, alternatief antwoord weer te geven. Als u gebruik maakt van die optie, wil u dan wel uw alternatieve antwoord eronder weergeven?

De overige zes vragen zijn open vragen. Geef zo goed mogelijk antwoord: het is niet fout als u iets niet weet. Ga in dat geval niet zo maar gokken, maar geef aan dat u het niet weet. Het telt immers niet mee voor de beoordeling van uw eindopdracht van de cursus.

### Vraag 1

Welk begrip hoort waar thuis?

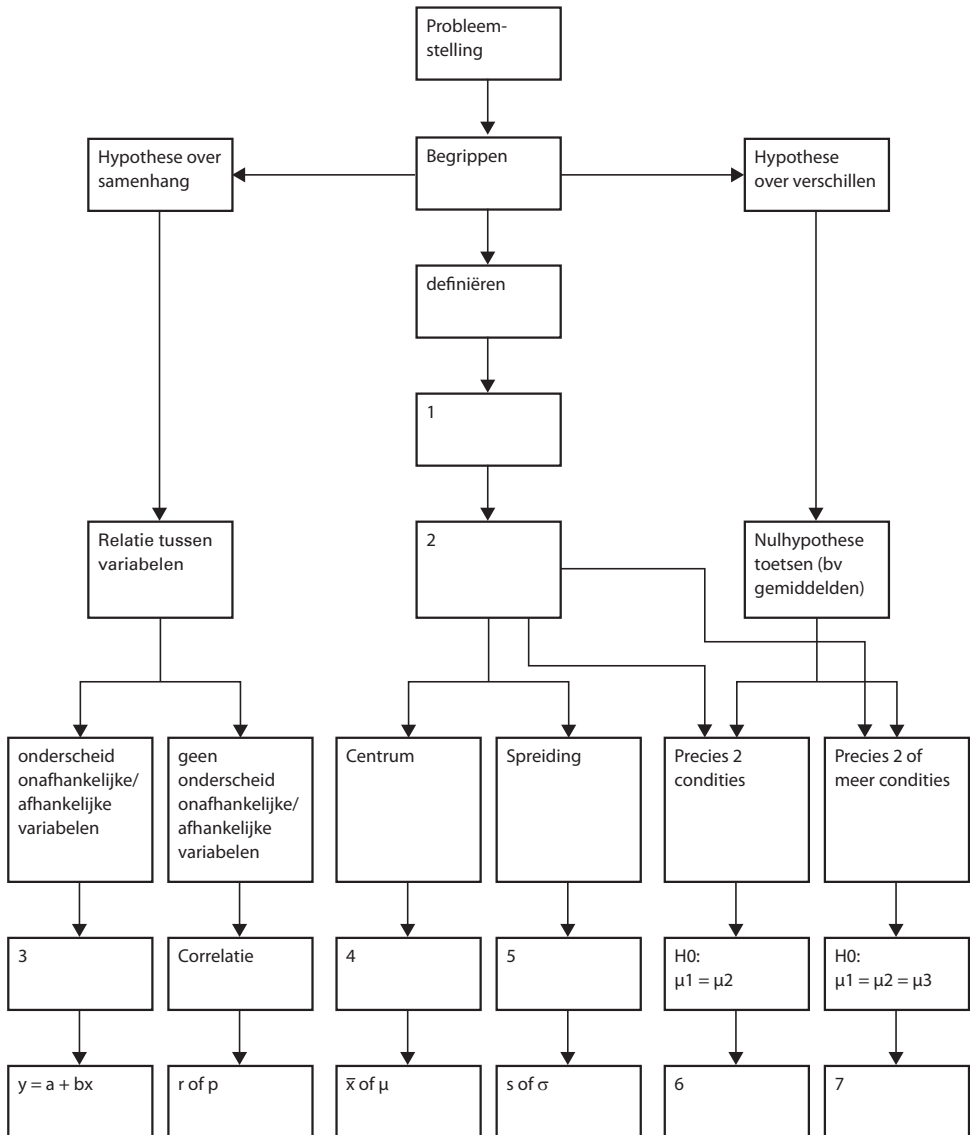
In het schema op de volgende bladzijde worden enkele begrippen uit onderzoek en statistiek met elkaar in verband gebracht. In 7 blokjes moeten nog begrippen worden ingevuld. In deze blokjes staat alleen een cijfer van 1 t/m 7. In het volgende lijstje vindt u een woord dat in de lege blokjes dient te staan. Vul in het lijstje achter een woord het cijfer in dat volgens u van toepassing is.

N.B. Een begrip mag meer dan een keer gebruikt worden.

N.B. Niet alle begrippen zijn van toepassing.



Woord	Cijfer
ANOVA	*****
Chi-kwadraat	*****
Correlatiecoëfficiënt	*****
Definiëren	*****
Experiment	*****
Geen onderscheid onafhankelijke en afhankelijke variabelen	*****
Gemiddelde	*****
Meetniveau	*****
Onderscheid afhankelijke en onafhankelijke variabelen	*****
Operationaliseren	*****
Regressie	*****
Relatie tussen variabelen	*****
T-toets	*****
Standaardafwijking	*****
Variabele	*****
Variantieanalyse	*****



**Vraag 2**

Een groep van 30 studenten hebben een toets gemaakt, bestaande uit 20 items. Het gemiddelde en de standaardafwijking van de toetsresultaten zijn berekend, de standaardafwijking bedraagt 0.

Op basis van deze gegevens weet u nu dat:

- a. ongeveer de helft van de scores boven het gemiddelde ligt.
- b. de test zo moeilijk is dat iedereen alle items fout had.
- c. er een fout in de berekening van de standaardafwijking moet hebben plaatsgevonden.
- d. iedereen hetzelfde aantal items correct heeft beantwoord.
- e. het gemiddelde, de mediaan en de modus van deze scores waarschijnlijk van elkaar verschillen.
- f. anders, namelijk ... (u kunt in de lege ruimte hieronder uw antwoord weergeven)

.....

.....

.....

**Vraag 3**

Een scriptiestudente heeft in haar onderzoek vastgesteld dat de relatie tussen verbale intelligentie en statistiekcijfers op een prestatietest in een steekproef van 45 vrouwelijke studenten .60 bedraagt en in een steekproef van 35 mannelijke studenten .70.

Probeer drie vragen (*in woorden en statistische hypothesevorm*) te bedenken waarop deze informatie een antwoord zou kunnen zijn.

Voorbeeld onderzoeksvraag 1:

.....

.....

.....

Voorbeeld onderzoeksvraag 2:

.....

.....

.....

.....

.....

.....

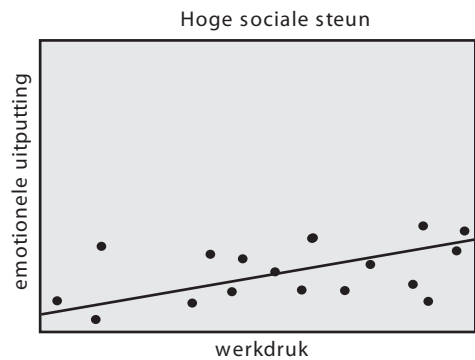
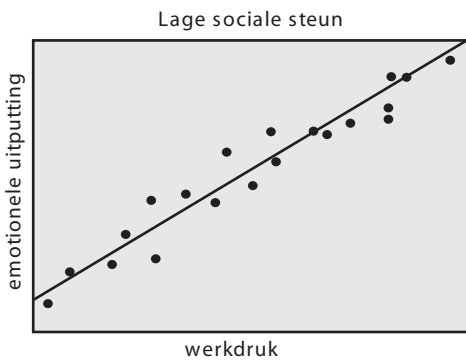
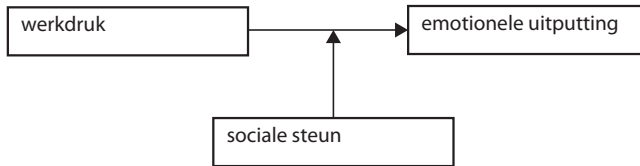
Zou u de relaties ook in een plaatje kunnen schetsen? Zo ja, schets dan het plaatje.

This image shows a full page of white paper with horizontal dashed lines, typical of primary-ruled notebook paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.



### Vraag 6

In de volgende twee spreidingsdiagrammen zijn de scores op emotionele uitputting en werkdruk bij een groep werknemers die veel sociale steun en bij een groep werknemers die weinig sociale steun in hun werkomgeving ondervinden afgebeeld (fictief).



Wat kunt u opmaken uit deze twee spreidingsdiagrammen over de relatie tussen werkdruk en emotionele uitputting? Geef uw mening zo uitgebreid mogelijk.

.....

.....

.....

.....

.....

.....

.....

.....

**Vraag 7**

Uit eerdere onderzoeken naar de relatie tussen ziekteverzuim en stress in zorginstellingen kwamen de volgende twee resultaten naar voren.

*Verpleegkundigen die op grote afdelingen werken, zijn gemiddeld per jaar veel vaker ziek dan verpleegkundigen die op kleine afdelingen werken.*

*Op grote afdelingen zijn verpleegkundigen met minder jaren werkervaring veel vaker ziek per jaar dan verpleegkundigen die veel jaren werkervaring hebben. Op kleine afdelingen is er geen verschil in ziekteverzuim tussen verpleegkundigen met veel of weinig werkervaring.*

De resultaten gaan gepaard met de volgende twee beleidsaanbevelingen:

1. *Beleidsadvies:* om het ziekteverzuim te verminderen zouden op grote afdelingen meer verpleegkundigen aangenomen moeten worden, zodat de gemiddelde werkdruk afneemt.
2. *Beleidsadvies:* Om het ziekteverzuim te verminderen is het aan te raden om verpleegkundigen die weinig ervaring hebben, op kleine afdelingen te laten werken en verpleegkundigen die veel jaren ervaring hebben, te laten werken op grote afdelingen.

Welk van de twee beleidsadviezen sluit het beste aan bij de onderzoeksbevinding?

- a. alleen advies 1
- b. alleen advies 2
- c. zowel advies 1 als advies 2
- d. géén van beide adviezen
- e. weet niet
- f. anders, namelijk

.....

.....

.....

.....

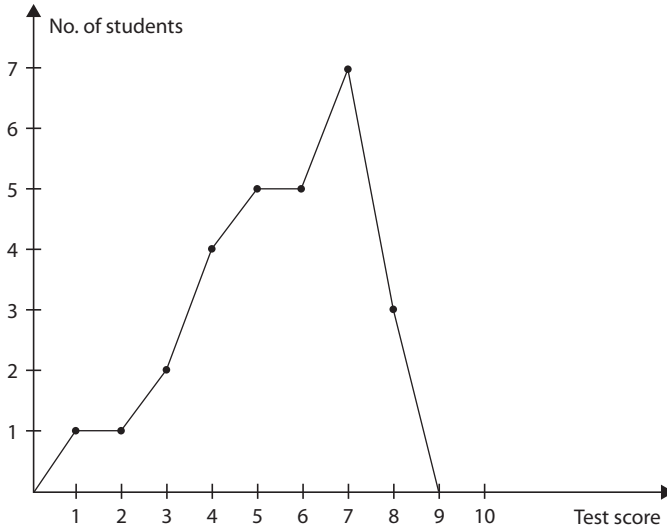
.....

.....

.....

### Vraag 8

Een onderzoeker is geïnteresseerd hoe studenten in een spellingtest presteerden om te kunnen besluiten of ze extra hulp zouden kunnen gebruiken. Onderstaande grafiek geeft de scores van 28 studenten op de spellingtest weer.



Als u zou worden gevraagd een getal of enkele getallen te noemen die het best de scores in de spellingtest van de 28 studenten zou weergeven, welk getal of welke getallen zou u dan kiezen?

Geef ook aan waarom en geef uw mening zo uitgebreid mogelijk.

.....

.....

.....

.....

.....

.....

.....

.....



**Vraag 9**

Het ontwikkelingscomité voor professionele leraren van de Unie van Leraren in Pennsylvania heeft een onderzoek in gang gezet om vast te stellen of het bezit van een Master certificaat tot een betere financiële beloning leidt van leraren aan het basisonderwijs. Leraren aan basisscholen, die als eindcertificaat hetzij het Bachelor-, hetzij het Master certificaat bezitten, uit de voorsteden van het schooldistrict, werden geselecteerd voor het onderzoek (dus onderwijzers die in het eerste geval geen hogere certificaten bezitten dan het Bachelorcertificaat en in het tweede geval niet hoger dan het Master certificaat). Alleen leraren met 10 tot 15 jaar ervaring maakten deel uit van de selectie. Per school werd er één leraar geselecteerd. Er werden geheel willekeurig (technisch: at random) 35 leraren geselecteerd met een Master certificaat en 35 leraren met een Bachelor certificaat. Informatie over de hoogte van hun jaarsalaris werd verkregen via de salarisadministraties, nadat daar toestemming voor was verkregen. Dit leidde tot de volgende resultaten:

	<i>Gemiddeld jaarsalaris</i>	<i>Standaarddeviatie</i>
Leraren met een Bachelor certificaat	\$ 32.100	\$ 5.100
Leraren met een Master certificaat	\$ 36.475	\$ 4.400

Als u kijkt naar de resultaten van het onderzoek, die hierboven zijn weergegeven, is er dan genoeg bewijs om te concluderen dat, gemiddeld gezien, basisschool-leraren met een Master certificaat meer verdienen dan basisschool-leraren met een Bachelor certificaat (gebruik  $\alpha = .05$ )?

.....

.....

.....

.....

.....

Methoden bij vraag 9:

Welke methoden heeft u bij vraag 9 gevolgd?

9a Type onderzoeksvraag:

- ☐ een samenvhangvraag
- ☐ een verschilvraag
- ☐ weet niet

9b Onderzoeksvariabelen:

- ☐ Het jaarsalaris en het type certificaat
- ☐ Het gemiddelde jaarsalaris van leraren met een Master en het gemiddelde jaarsalaris van leraren met een Bachelorcertificaat
- ☐ Het jaarsalaris, het type certificaat, het aantal jaren onderwijservaring
- ☐ Het jaarsalaris, het type certificaat, het aantal jaren onderwijservaring en locatie (Centrum Pennsylvanië of voorsteden)
- ☐ Weet niet

9c Wat is het meetniveau van de afhankelijke variabele?

- ☐ interval
- ☐ nominaal
- ☐ ordinaal
- ☐ ratio
- ☐ weet niet

9d Welke statistische procedure zou u gebruiken om tot uw conclusie te komen?

- ☐ chi-kwadraat toets
- ☐ correlatie-analyse
- ☐ t-toets
- ☐ regressie-analyse
- ☐ variantie-analyse
- ☐ weet niet

9e Aan welke statistische voorwaarde(n) moet zijn voldaan om voorgaande statistische techniek te mogen uitvoeren?

- ☐ Géén voorwaarden nodig
- ☐ Eén voorwaarde: De populatieverdeling waaruit steekproef afkomstig is moet normaalverdeeld zijn.
- ☐ Eén voorwaarde: De varianties van de populatieverdelingen waaruit steekproef afkomstig is moeten gelijk zijn.
- ☐ Twee voorwaarden: De populatieverdeling waaruit elke groep afkomstig is moet normaalverdeeld zijn én de varianties gelijk.
- ☐ Weet niet

9f Eventueel toelichting bij voorgaande keuzen.

.....

.....

.....

.....

**Affectieve factoren t.o.v. statistiek**

Welke ideeën leven bij u over het vak statistiek?

Geef bij de volgende stellingen aan in hoeverre u er het het oneens of eens bent. U heeft de keuzemogelijkheid uit vijf antwoorden per item, variërend van 1 = 'geheel oneens' tot 5 = 'geheel mee eens'. Omcirkel het antwoord dat het meest overeenstemt met uw eigen mening en denk s.v.p. niet te lang na over uw antwoord.

		geheel oneens	oneens	oneens noch eens	mee eens	geheel mee eens
146	Ik ben graag bezig met statistiek.	1	2	3	4	5
147	Ik voel me onzeker wanneer ik statistische problemen moet oplossen.	1	2	3	4	5
148	Mijn manier van denken zorgt ervoor dat ik moeite heb om statistiek te begrijpen.	1	2	3	4	5
149	Statistische formules zijn gemakkelijk te begrijpen.	1	2	3	4	5
150	Statistiek is waardeloos.	1	2	3	4	5
151	Statistiek is een ingewikkeld vak.	1	2	3	4	5
152	Statistiek zou een verplicht onderdeel moeten zijn van elke academische opleiding.	1	2	3	4	5
153	Statistische vaardigheden zullen mij gemakkelijker aan een baan helpen.	1	2	3	4	5
154	Ik heb er geen idee van wat er in deze statistiekcursus gebeurt.	1	2	3	4	5
155	Statistiek is niet nuttig voor de gemiddelde academicus.	1	2	3	4	5
156	Statistiek tentamens frustreren mij.	1	2	3	4	5
157	Statistisch denken is niet van toepassing in mijn leven buiten mijn werk.	1	2	3	4	5
158	Ik gebruik statistiek in het leven van alledag.	1	2	3	4	5
159	Ik ben gestrest tijdens bijeenkomsten statistiek.	1	2	3	4	5

		geheel oneens	oneens	oneens noch eens	mee eens	geheel mee eens
160	Ik volg graag cursussen statistiek.	1	2	3	4	5
161	Statistische conclusies worden zelden getrokken in het dagelijkse leven.	1	2	3	4	5
162	Statistiek is een vak dat de meeste mensen snel leren.	1	2	3	4	5
163	Statistiek leren vereist heel wat discipline.	1	2	3	4	5
164	Ik zal statistiek niet toepassen in mijn beroep.	1	2	3	4	5
165	Ik maak heel wat wiskundefouten in statistiek.	1	2	3	4	5
166	Ik ben bang van statistiek.	1	2	3	4	5
167	Statistiek betekent veel rekenwerk.	1	2	3	4	5
168	Ik kan statistiek leren.	1	2	3	4	5
169	Ik begrijp statistische vergelijkingen.	1	2	3	4	5
170	Statistiek is irrelevant in mijn leven.	1	2	3	4	5
171	Statistiek is uiterst technisch.	1	2	3	4	5
172	Ik vind het moeilijk om statistische concepten te begrijpen.	1	2	3	4	5
173	De meeste mensen moeten een nieuwe manier van denken leren om zich statistiek eigen te maken.	1	2	3	4	5

### Algemeen

En als laatste... Heeft u nog op- of aanmerkingen over de onderzoeksmodules, dan horen we dat graag van u.

Bedankt voor uw medewerking.

.....

.....

.....

.....

.....

.....



## **Vragenlijst evaluatie T<sub>1</sub>-meting**

**Vragenlijst voorkennis van en mening over leerstof methoden en technieken van onderzoek en statistiek**

Geachte student,

Allerlei ideeën en meningen leven over cursussen Methodenleer en Statistiek. Ze worden weliswaar door opleidingen van belang geacht om psychologie, bedrijfskunde, bestuurskunde, sociologie, pedagogiek, onderwijskunde, enzovoorts te kunnen studeren, maar het is ook algemeen bekend dat deze cursussen voor veel studenten een struikelblok vormen. Is daar nu niets aan te doen? Een mogelijkheid is om de didactiek aan te passen. In een aantal experimenten wordt daar nu ervaring mee opgedaan. De cursus die thans voor u ligt, is daar een product van. In deze cursus wordt de leerstof over de methodenleer en de statistiek op een andere manier aangeboden dan u wellicht gewend bent. Maar werkt zo'n verandering in didactiek ook? Dat is de vraag die we ons gesteld hebben. Daarom moeten we weten welke ideeën er over deze vakken leven, hoe u nú denkt over deze vakken en hoe dat over een tijdje zal zijn als u deze cursus hebt afgerond. Alleen dan kunnen we inspelen op de behoeften van onze studenten, zonder aan kwaliteit in te boeten. Daarom vragen we u deze vragenlijst aan het begin van de cursus in te vullen. Opdat we het effect van dit nieuw soort onderwijs op uw kennis, inzichten en welbevinden kunnen nagaan, is het nodig dat we u in deze cursus kunnen 'volgen'. Daarvoor is uw studentnummer nodig. Wij verzoeken u deze te willen invullen. Uiteraard worden uw gegevens zeer vertrouwelijk behandeld. Wij danken u bij voorbaat voor uw medewerking.

U kunt de ingevulde lijst in een envelop zonder postzegels terugsturen naar het volgende adres:

Open Universiteit Nederland  
Directoraat Onderwijs – Psychologie  
t.a.v. drs. J.A. van Buuren  
Antwoordnummer 4400  
6400 VC Heerlen

Succes met uw studie!

Met vriendelijke groet,

Hans van Buuren.

- I Wat is uw studentnummer? 83.....  
 N.B. Uw studentnummer bestaat uit 9 cijfers. Probeer deze foutloos over te nemen.
- II Wat is uw postcode: .....
- III Wat is uw huisnummer? .....


Hieronder staan een aantal vragen over allerlei aspecten die te maken hebben met het doen van onderzoek en de statistische verwerking van gegevens. We willen door middel van deze vragenlijst te weten komen wat u *op dit moment* weet over onderzoek en statistische analyse. Deze informatie kunnen we gebruiken om achteraf na te gaan of de opdrachten uit de cursus niet te moeilijk of te makkelijk zijn.

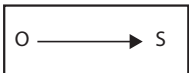
We vragen u deze lijst zorgvuldig in te vullen. Lees de vraag eerst rustig door en geef daarna antwoord. Bij de meeste vragen kunt u een van de antwoordalternatieven kiezen door uw keuze te omcirkelen. Er zijn enkele open vragen bij. Geef zo goed mogelijk antwoord: het is niet erg als u iets niet weet. Ga in dat geval niet zomaar gokken, maar geef aan dat u het niet weet. Het telt immers niet mee voor de beoordeling van uw eindopdracht van de cursus.


1. Op het terrein van onderzoek worden veel specifieke termen gebruikt. Hieronder staat een lijst met enkele onderzoekstermen. We willen graag voor elke term weten of u de term en de betekenis er van kent. U kunt kiezen uit de volgende drie antwoorden:
1. ik ken de term en weet wat het betekent;
  2. ik ken wel de term, maar weet niet precies wat het betekent;
  3. ik heb nog nooit van deze term gehoord.


frequentie	1	2	3
modus	1	2	3
variabele	1	2	3
data	1	2	3
analysemethode	1	2	3
gemiddelde	1	2	3
samenhang (tussen variabelen)	1	2	3
kruistabel	1	2	3
onafhankelijke variabele	1	2	3
standaardafwijking of -deviatie	1	2	3
aselecte steekproef	1	2	3
normaalverdeling	1	2	3
causale relatie	1	2	3
significantieniveau	1	2	3
regressie-analyse	1	2	3
correlatiecoëfficiënt	1	2	3
meetniveau	1	2	3
verklaarde variantie	1	2	3



2. Welke van de onderstaande termen is een *variabele*?
- de universiteit
  - de Open Universiteit Nederland
  - de studenten van de Open Universiteit Nederland
  - de leeftijd van de studenten van de Open Universiteit Nederland
  - weet niet
3. Hoeveel *variabelen* zitten in de volgende uitspraak?  
 “Hoe sterker de sociale controle in een wijk, des te geringer is het aantal inbraken in die wijk”.
- 1
  - 2
  - 3
  - 4
  - weet niet
4. Uit een onderzoek komt het volgende resultaat naar voren: iemands opleidingsniveau hangt positief samen met het aantal keren dat zij of hij een schouwburgvoorstelling bezoekt. Welke van de onderstaande conclusies is onderzoekstechnisch juist?
- het opleidingsniveau is de oorzaak van het aantal keren dat iemand een schouwburgvoorstelling bezoekt
  - het aantal keren dat iemand een schouwburgvoorstelling bezoekt, is van invloed op haar of zijn opleidingsniveau
  - naarmate iemand een hogere opleiding heeft, bezoekt zij of hij vaker een schouwburgvoorstelling
  - naarmate iemand een hogere opleiding heeft, waardeert zij of hij het gebodene positiever.
  - weet niet
5. Welke van de volgende schema's, geeft de samenhang uit de vorige vraag tussen opleidingsniveau (O) en het aantal bezoeken van een schouwburgvoorstelling (S) correct weer.
- a 

b 

c 

d 
- weet niet

6. Noem drie manieren om onderzoeksgegevens te verzamelen:
- a .....
- b .....
- c .....
7. U krijgt de opdracht om te onderzoeken of de mate van ziekteverzuim onder bejaardenverzorgers te maken heeft met de stress die de bejaardenverzorgers ervaren. Het is de bedoeling dat u interviews houdt.
- A Bij welke groep mensen zou u het beste de interviews kunnen afnemen? (Slechts één antwoord omcirkelen)
- a bejaardenverzorgers
- b managementleden van verschillende zorginstellingen
- c bejaarden
- d geen van deze groepen
- e weet niet
- B Noem drie onderwerpen die u in de interviews aan de orde zou stellen:
- 1 .....
- 2 .....
- 3 .....
8. U wilt graag inzicht krijgen in de mate waarin jongeren tussen 18 en 23 jaar uitgaan.
- A Op de website van een horeca-instelling, vindt u de volgende tabel.

<i>Aantal keren per maand dat jongeren tussen 18 en 23 jaar een discotheek bezoeken</i>	<i>Aantal jongeren</i>
méér dan 4 keer per maand	20
1 á 4 keer per maand	120
minder dan 1 keer per maand	60
Totaal aantal jongeren	200

Hoe groot is het *percentage* jongeren dat méér dan 4 keer per maand een discotheek bezoekt?

- a 5 %
- b 10 %
- c 20 %
- d 40 %
- e weet niet

B Op dezelfde site komt u een soortgelijke tabel tegen, maar is er onderscheid gemaakt tussen jongens en meisjes.

<i>Aantal keren per maand dat jongeren tussen 18 en 23 jaar een discotheek bezoeken</i>	<i>Aantal jongens</i>	<i>Aantal meisjes</i>	<i>Totaal aantal</i>
méér dan 4 keer per maand	15	5	20
1 á 4 keer per maand	90	30	120
minder dan 1 keer per maand	45	15	60
Totaal aantal jongeren	150	50	200

Welke van de volgende uitspraken is juist?

- a Jongens gaan verhoudingsgewijs vaker ‘méér dan 4 keer per maand’ naar een discotheek dan meisjes.
  - b Jongens gaan verhoudingsgewijs even vaak ‘1 á 4 keer per maand’ naar een discotheek als meisjes.
  - c Meisjes gaan verhoudingsgewijs vaker ‘minder dan 1 keer per maand’ naar een discotheek dan jongens.
  - d weet niet
9. Uit eerdere onderzoeken naar de relatie tussen ziekteverzuim en stress in zorginstellingen kwamen de volgende twee resultaten naar voren.
- Verpleegkundigen die op grote afdelingen werken zijn gemiddeld per jaar veel vaker ziek, dan verpleegkundigen die op kleine afdelingen werken.
  - Op grote afdelingen zijn verpleegkundigen met minder jaren werkervaring veel vaker ziek per jaar, dan verpleegkundigen die veel jaren werkervaring hebben. Op kleine afdelingen is er geen verschil in ziekteverzuim tussen verpleegkundigen met veel of weinig werkervaring.

De resultaten gaan gepaard met de volgende twee beleidsaanbevelingen:

- 1 *Beleidsadvies:* om het ziekteverzuim te verminderen zouden op grote afdelingen meer verpleegkundigen aangenomen moeten worden, zodat de gemiddelde werkdruk afneemt.
- 2 *Beleidsadvies:* Om het ziekteverzuim te verminderen is het aan te raden om verpleegkundigen die weinig ervaring hebben op kleine afdelingen te laten werken en verpleegkundigen die veel jaren ervaring hebben te laten werken op grote afdelingen.

Welke van de twee beleidsadviezen sluiten het beste aan bij de onderzoeksbevinding.

- a Alleen advies 1
  - b Alleen advies 2
  - c Zowel advies 1 als advies 2
  - d Géén van beide adviezen
  - e weet niet
10. De directie van een opleiding signaleert het probleem dat studenten niet tevreden zijn over het onderwijs dat ze krijgen. Dit blijkt telkens uit de evaluaties die studenten na een module moeten invullen. Een onderzoeker wil nagaan waarover de studenten nu niet tevreden zijn en wat de reden daarvan is. Zij besluit om de docenten van verschillende modules te onderwerpen over hun ervaringen met de studenten in de klas. Haar motivatie waarom ze docenten interviewt, is dat volgens haar docenten weten wat er onder de studenten leeft en ze ze vrijwel elke dag in de klas meemaken.

Vindt u het een logische keuze om docenten te interviewen?  
Motiveer uw antwoord.

.....

.....

.....

.....

.....

.....

.....

.....

.....

11. In een discussiegroep stelt iemand dat ‘vrouwen niet kunnen autorijden’.  
 Op welke manier zout u – zonder u aangesproken te voelen, beledigd te zijn of van harte er mee in te stemmen – op een ‘onderzoekstechnische’ manier de steller van deze uitspraak van repliek dienen? Welke ‘inhoudelijke’ vragen zou u hem of haar willen stellen? Motiveer uw antwoord.  
 Zou u er ook een statistisch (kruis)tabelletje bij kunnen tekenen?  
 Zo ja, teken het dan.

.....

.....

.....

.....

.....

.....

.....

.....

12. Kent u de statistische termen *gemiddelde*, *standaardafwijking* (ook wel standaarddeviatie genoemd) en *normaalverdeling* en zo ja, wilt u dan kort omschrijven wat u daaronder verstaat. Beschrijf een bekende vuistregel in eigen woorden. Zou u de relaties ook in een plaatje kunnen schetsen? Zo ja, schets dan het plaatje.

.....

.....

.....

.....

.....

.....

.....

13. *Voorbeeld:* Een onderzoeker heeft een aantal eerstejaars psychologiestudenten geïnterviewd en geconstateerd dat het merendeel (73%) liever een boek bestudeert en daarover een meerkeuzevragentamen aflegt dan dat ze daarover een werkstuk met een eigen probleemomschrijving moeten schrijven.

*Vraag:* Kunt u aangeven aan welke voorwaarden de onderzoekssituatie moet voldoen, of korter gezegd welke statistische begrippen eraan te pas komen om deze bevinding te kunnen generaliseren naar een grotere groep studenten dan alleen de geïnterviewde groep psychologiestudenten? Zo ja, motiveer uw antwoord door de betreffende statistische termen zoveel mogelijk puntsgewijs te noemen.

.....

.....

.....

.....

.....

.....

.....

.....

In het volgende deel van de vragenlijst willen we nagaan welke ideeën bij u leven over de cursussen onderzoeksmethoden en de statistiek.

Wilt u in de volgende lijst aangeven in hoeverre u het oneens of eens bent met elk van de volgende uitspraken door steeds het vakje aan te kruisen dat het best uw mening weergeeft?

De antwoordmogelijkheden zijn:

- helemaal oneens
- gedeeltelijk eens
- oneens noch eens
- gedeeltelijk mee eens
- helemaal mee eens

		hele- maal oneens	gedeel- telijk oneens	oneens noch eens	gedeel- telijk mee eens	hele- maal mee eens
1	Ik geef de voorkeur aan uitdagende taken en opdrachten zodat ik nieuwe dingen leer.					
2	Vergeleken met andere studenten methodenleer en statistiek verwacht ik het beter te zullen doen.					
3	Ik ben zo gespannen tijdens een tentamen dat ik me de dingen die ik heb geleerd niet kan herinneren.					
4	Statistiek kan ik te allen tijde gebruiken.					
5	Niet elke student die psychologie studeert kan statistiek leren.					
6	Bij de beroepsuitoefening van de gemiddelde psycholoog is kennis van statistiek niet vereist.					
7	Ik zou liever een psychologieopleiding zien zonder statistiek.					
8	Ik heb statistiek nooit graag gedaan en het is voor mij een van de meest gevreesde vakken.					
9	Ik zou graag mijn onderzoekstechnische en statistische vaardigheden verder ontwikkelen.					
10	Statistiek is in de psychologie niet zo waardevol en zeker niet noodzakelijk.					
11	Ik denk dat ik na afloop van een eerste cursus Statistiek in staat ben een eenvoudig onderzoek statistisch te verwerken.					

		hele- maal oneens	gedeel- telijk oneens	oneens noch eens	gedeel- telijk mee eens	hele- maal mee eens
12	Oefeningen voor statistiek los ik op een automatische manier op, zonder inzicht in de leerstof.					
13	Ik maak nooit meer statistieksommen dan zijn opgegeven in de cursus.					
14	Ik heb niet de bedoeling en ik ben ook niet bereid na een verplichte cursus statistiek nog verder vrijwillig een statistiekcursus te gaan studeren.					
15	Vrijwel elke student kan statistiek leren als het op een gepaste manier wordt onderwezen.					
16	Statistiek is saai en vervelend omdat het geen ruimte laat voor een persoonlijke visie.					
17	Statistiek vind ik prettig en stimulerend.					
18	Dikwijls als ik met getallen, formules en tabellen moet werken, voel ik me onzeker en slecht op mijn gemak.					
19	Ik heb de bedoeling en ik ben ook bereid statistiek buiten het kader van een cursus statistiek te gebruiken.					
20	Zodra ik met een statistisch probleem begin, is het net alsof mijn denken verstart, geblokkeerd wordt.					
21	Statistiek kan niet voor iedere student begrijpelijk gemaakt worden.					
22	Ik heb geen moeilijkheden met het gebruik en toepassen van begrippen, symbolen, formules en tabellen in de statistiek.					
23	Direct na afloop van een les of een college ben ik niet in staat tentamen af te leggen over de bestudeerde leerstof.					
24	Ik werk graag met cijfers, formules, symbolen, tabellen, enzovoorts.					
25	Indien een medestudent mij uitleg vraagt over een of ander onderdeel van de statistiek, ben ik meestal niet in staat om die uitleg te geven.					
26	Statistiek is zeer interessant en ik heb meestal genoeg beleefd aan de cursus(sen) over dit onderwerp.					
27	Statistiek stelt je in staat bepaalde begrippen met meer zekerheid te gebruiken.					
28	Ik denk dat ik het na afloop van een cursus statistiek niet zal aandurven een eenvoudig onderzoek statistisch te verwerken					



		hele- maal oneens	gedeel- telijk oneens	oneens noch eens	gedeel- telijk mee eens	hele- maal mee eens
29	Ik heb steeds graag statistiek gestudeerd.					
30	Statistiek maakt me nerveus en ongemakkelijk.					
31	Ik geloof dat ik geen goede psycholoog kan worden zonder iets van statistiek af te weten.					
32	Kennis van de statistiek is noodzakelijk om met succes de psychologieopleiding te kunnen volgen.					
33	Indien een medestudent mij uitleg vraagt over een of ander onderdeel van de statistiek, dan durf ik het wel aan die uitleg te geven.					
34	Nadat ik oefeningen heb gemaakt voor statistiek, begrijp ik de leerstof veel beter.					
35	Er is niets creatiefs in de statistiek; het is enkel en alleen het toepassen van begrippen en formules.					
36	Het begrijpen en kennen van de statistiek is zowel nodig voor de onderzoekspsycholoog als voor de practicus.					
37	Direct na afloop van een oefencollege zou ik tentamen over de bestudeerde leerstof durven af te leggen.					
38	Statistiekopgaven zijn steeds zeer moeilijk voor mij.					
39	Statistiek is niet belangrijk in het dagelijkse leven.					
40	Ik los liever een statistieksom op dan een aantal zaken uit het hoofd te moeten leren.					
41	Iedereen met een gemiddelde intelligentie kan statistiek leren.					
42	Ik vind het belangrijk om onderzoeksmethoden en statistiek te leren.					
43	Ik leer graag iets over onderzoeksmethoden en statistiek.					
44	Ik ben er zeker van dat ik deze cursus goed zal kunnen volgen.					
45	Ik denk dat ik dat wat ik hier over onderzoeksmethoden en statistiek leer, kan gebruiken voor andere vakken.					
46	Ik verwacht dat ik deze cursus met heel goede resultaten zal afronden.					
47	Ik denk dat ik een goede student ben in vergelijking met andere studenten die methodenleer en statistiek moeten doen.					

		hele- maal oneens	gedeel- telijk oneens	oneens noch eens	gedeel- telijk mee eens	hele- maal mee eens
48	Ik zou die onderwerpen voor het maken van een werkstuk kiezen waarvan ik veel kan leren, ook al betekent dit meer werk.					
49	Ik ben er zeker van dat ik de taken en opdrachten van deze cursus heel goed zal kunnen maken.					
50	Ik heb een ongemakkelijk en ellendig gevoel als ik tentamen doe.					
51	Ik denk dat ik een hoog cijfer voor de eindopdracht van deze cursus zal halen.					
52	Ik probeer van mijn fouten te leren, zelfs al presteer ik goed op een tentamen.					
53	Ik denk dat de stof die ik in deze cursus zal leren belangrijk voor me is om te leren.					
54	Mijn studievoordigheden zijn uitstekend vergeleken met die van andere studenten die methodenleer en statistiek moeten doen.					
55	Ik denk dat ik interessante dingen ga leren in deze cursus.					
56	Ik denk dat ik in staat zal zijn het materiaal van deze cursus te begrijpen.					
57	Ik maak me zorgen over tentamens.					
58	Begrip van statistische procedures is belangrijk voor mij.					
59	Terwijl ik tentamen doe, denk ik eraan hoe slecht ik het doe.					
60	Als ik me voorbereid op een tentamen, probeer ik informatie uit verschillende bronnen te integreren.					
61	Tijdens het studeren probeer ik mij datgene wat op de bijeenkomsten is gezegd te herinneren zodat ik de werkboek-opdrachten zo goed mogelijk kan uitvoeren.					
62	Ik stel mezelf vragen om er zeker van te zijn dat ik het materiaal dat ik heb bestudeerd beheers.					
63	Ik vind het moeilijk om de hoofdpunten te ontdekken in hetgeen ik lees.					
64	Als het studiemateriaal moeilijk is geef ik het òf op òf ik bestudeer alleen de eenvoudige delen.					
65	Tijdens het studeren probeer ik belangrijke kwesties in mijn eigen woorden te formuleren.					

		hele- maal oneens	gedeel- telijk oneens	oneens noch eens	gedeel- telijk mee eens	hele- maal mee eens
66	Tijdens het studeren probeer ik altijd te begrijpen wat er precies bedoeld wordt, zelfs al is het niet direct van belang.					
67	Als ik een tentamen voorbereid probeer ik mezelf zoveel mogelijk dingen te herinneren.					
68	Om mezelf te helpen het materiaal te onthouden lees ik mijn aantekeningen telkens over.					
69	Als een opdracht niet verplicht is, dan maak ik toch die opdracht.					
70	Ik werk door totdat ik klaar ben, zelfs al is het studiemateriaal saai en vervelend.					
71	Als ik me voorbereid op een tentamen, zeg ik de te leren feiten voor mezelf steeds hardop.					
72	Voordat ik met studeren begin, denk ik na over de dingen die ik wil gaan doornemen.					
73	Bij het maken van nieuwe opdrachten maak ik gebruik van hetgeen ik heb geleerd in andere opdrachten.					
74	Als ik één onderwerp bestudeer, probeer ik materiaal van verschillende bronnen bij elkaar te brengen.					
75	Terwijl ik aan het studeren ben stop ik een enkele keer, om daarna de draad weer op te pakken.					
76	Als hulp bij het studeren maak ik een samenvatting van de kernpunten uit het studiemateriaal.					
77	Ik werk hard om een hoog cijfer te halen, zelfs al vind ik een cursus niet leuk.					
78	Tijdens het studeren probeer ik de dingen die ik lees te verbinden met datgene wat ik reeds weet					
79	Wiskunde boeit me.					
80	Ik voel me onzeker in een groepje mensen waar wiskunde het onderwerp is.					
81	Ik probeer wiskunde altijd te vermijden.					
82	Ik ben van nature goed in wiskunde.					
83	Ik ben niet goed in cursussen waar ik moet rekenen.					
84	Ik hou van wiskunde omdat het zo exact is.					
85	Voor statistische vraagstukken moet je bij mij zijn.					

		hele- maal oneens	gedeel- telijk oneens	oneens noch eens	gedeel- telijk mee eens	hele- maal mee eens
86	Ik beheers een aanzienlijk deel van de statistische leerstof die voor deze cursus nodig is.					
87	Het maakt niet uit hoe hard ik ook studeer, ik zal het nooit goed doen in een discussiegroep als het over statistiek gaat.					

Wilt u bij onderstaande vragen aangeven wat uw mening is over statistiek.

- 88 In hoeverre vindt u kennis van de statistiek nuttig?  
Beargumenteer uw antwoord.

- 89 Denkt u dat een statistisch computerpakket als SPSS (meegeleverd met de cursus) een verplichte statistiekcursus zou kunnen vergemakkelijken?  
Beargumenteer uw antwoord.

- 90 Welke onderwerpen of onderzoeksaspecten vindt u interessant binnen de statistiek?  
Beargumenteer uw antwoord.

- 
- 91 Als u vrijwillig een statistiekcursus zou gaan volgen, waarom denkt u dat deze dan nuttig zou zijn? Beargumenteer uw antwoord.

- 
- 92 Welke plaats neemt statistiek in binnen uw studie? Beargumenteer uw antwoord.

- 
- 93 In hoeverre vindt u statistiek bedreigend en angstaanjagend? Beargumenteer uw antwoord.

- 
- 94 In hoeverre vindt u statistiek boeiend? Beargumenteer uw antwoord.
-

Ten slotte willen we u graag een vraag stellen over uw ervaring met computers.  
Hoe vaak maakt u gemiddeld gebruik van de volgende programma's?

	<i>meer dan 1x per week</i>	<i>1x per week</i>	<i>1x per 2 weken</i>	<i>1x per maand</i>	<i>minder dan 1x per maand</i>	<i>nooit</i>
- Internet						
- E-mail						
- Tekstverwerkingsprogramma's (bijv. Word, WP)						
- Andere programma's, te weten						
1 .....						
2 .....						
3 .....						
4 .....						
5 .....						

Wij danken u voor de tijd die u genomen hebt om de vragenlijst in te vullen. Hebt u opmerkingen over deze vragenlijst of deze manier van onderzoeken, dan zouden wij het op prijs stellen indien u deze hierna wilt opschrijven.



## **Vervolgonderzoek naar nonrespons in het crosssectionele onderzoek**



## Abstract

Nonrespons is een veelvoorkomend probleem bij surveyonderzoek. In de crosssectionele evaluatie van dit proefschriftonderzoek bedraagt de nonrespons bijna 80%. De vraag is of een respons van 20% voldoende is om generaliserende uitspraken te doen met betrekking tot de resultaten. In een vervolgonderzoek naar de nonrespons is aan een aselechte steekproef van de nonrespondenten een verkorte vragenlijst voorgelegd. Aangezien de meetinstrumenten uit de oorspronkelijke vragenlijst onderworpen zijn aan Raschanalyses was het mogelijk om met één of enkele items een directe vergelijking tussen respondenten en nonrespondenten te maken. In de analyses zijn twee groepen respondenten te onderscheiden naar de wijze waarop deze het statistiekonderwijs hebben gevolgd, te weten: vakgericht (VAK) of competentiegericht (COMP). Gebleken is dat de nonrespons is toe te schrijven aan situationele factoren die niets met het onderwerp van onderzoek van doen hebben. Er zijn géén aanwijzingen gevonden dat de lage nonrespons tot vertekening ('nonresponsbias') heeft geleid.

## Inleiding

Nonrespons is geen nieuw verschijnsel bij surveyonderzoek en elke onderzoeker krijgt er mee te maken. De vraag dient zich dan ook aan wat een acceptabel responsniveau is. Zo wordt vaak gebruik gemaakt van de responsregel van Babbie (2007, p. 262) die stelt dat 50% respons adequaat is, 60% goed en 70% zeer goed. In de meeste gevallen gaat men er echter van uit dat een respons van 70% tot 80% noodzakelijk is om generaliserende uitspraken te kunnen doen. Dat nonrespons een probleem kan zijn, wordt door wetenschappers niet ontkend. Zo stellen Viswesvaran, Barrick en Ones in 1993 al dat nonrespons een bedreiging vormt voor de externe validiteit van conclusies uit surveys. Er wordt echter nauwelijks serieuze aandacht besteed aan de vraag wanneer nonrespons een bedreiging vormt en hoe hiermee dient te worden omgegaan. Door de socioloog Ellemers (2004) is dit betiteld als het 'schandaal van de non-respons', waarbij jarenlang is ontkend of gebagatelliseerd dat de uitkomsten van een surveyonderzoek zelfs bij een nonrespons van 20% of 30% al dubieus kunnen zijn. Dit dubieuze karakter heeft vooral te maken met het feit dat er geen informatie is over de verdeling van de nonrespondenten met betrekking tot de afhankelijke variabelen. Wetenschappers hanteren hiervoor het standaardantwoord dat 'als de steekproef maar groot genoeg is, er niets aan de hand is omdat het slechts gaat om verbanden tussen variabelen en die blijven dezelfde ongeacht de omvang van de nonrespons', maar dat is onzin, aldus Ellemers.

Deze visie van Ellemers wordt ondersteund door Stoop (2005) die in haar proefschrift concludeert dat het niet de omvang van de groep nonrespondenten is die zorgt voor een vertekening van de resultaten, de zogenaamde nonresponsbias, maar juist de kenmerken van deze groep. Is er namelijk sprake van MCAR-nonrespons ('missing completely at random'), dan leidt dit niet tot vertekening. Sterker nog, in een dergelijk geval kan een extra inspanning om een hogere respons te verkrijgen zelfs leiden tot meer vertekening.

Volgens Stoop (2005) kunnen voor nonrespons drie relatief onafhankelijke oorzaken worden aangewezen, te weten: onbereikbaarheid, onvermogen en onbereidwilligheid. Er is sprake van onbereikbaarheid als mensen niet veel thuis zijn, de telefoon niet opnemen of als e-mail niet door het spamfilter komt. Van onvermogen wordt gesproken als mensen niet mee kunnen doen, bijvoorbeeld omdat ze ziek zijn, niet kunnen lezen of de taal slecht begrijpen en/of spreken. Bij onbereidwilligheid ten slotte gaat het om mensen die wel bereikt zijn en wel kunnen meewerken maar die dat weigeren om wat voor reden dan ook. Omdat 'onvermogen' veel minder voorkomt dan 'onbereikbaarheid' en 'onbereidwilligheid', gaat in de literatuur de meeste aandacht uit naar het onderscheid tussen 'bereikbaarheid' en 'bereidwilligheid'. Als er wordt gesproken over 'bereikbaarheid' gaat het in de praktijk vooral om face-to-face of telefonisch onderzoek.

Zowel voor 'onbereikbaarheid' als voor 'onbereidwilligheid' is getracht oorzaken te definiëren. Bij het onderzoek naar bereikbaarheid bijvoorbeeld blijkt dat het tijdstip waarop geprobeerd wordt contact te leggen van belang is ('s avonds is de kans op succes om voor de hand liggende redenen groter dan overdag). Contactpogingen op verschillende tijdstippen over een langere periode leiden tot een zeer hoog contactpercentage. Voor de studentenpopulatie van de Open Universiteit Nederland geldt dat velen een e-mailadres hebben en in elk geval een adres. Bereikbaarheid speelt dus niet voor deze populatie. Om die reden is bereikbaarheid niet van belang en gaat het vooral om bereidwilligheid. De oorzaken van 'onbereidwilligheid' zijn complex en hangen vaak onderling samen waardoor het moeilijk is een bepalende factor te identificeren. Zo worden in de analyse van nonrespons vaak verschillende leeftijdsgroepen onderscheiden. Groves en Couper (1998) vonden echter geen eenduidig effect voor leeftijd. In sommige gevallen participeerden ouderen minder en in sommige gevallen meer. Een verklaring hiervoor zou kunnen zijn dat niet leeftijd maar een daarmee samenhangende variabele voor het effect verantwoordelijk was. Zo vindt men onder ouderen bijvoorbeeld veel alleenstaande vrouwen met een lage opleiding. Naar de feitelijke oorzaak van de 'onbereidwilligheid' (leeftijd, sekse, burgerlijke staat, opleidingsniveau) kan dan alleen maar geraden worden.

Er zijn verschillende manieren om de invloed van nonrespons te reduceren. Als er relevante oorzaken voor '(on)bereidwilligheid' aan een bepaald surveyonderzoek bekend zijn, dan kan daar in de analyses door middel van weging rekening mee worden gehouden. Stel dat bij een onderzoek naar theaterbezoek blijkt dat er nauwelijks alleenstaanden hebben meegedaan en dat uit de literatuur blijkt dat alleenstaanden niet vaak naar het theater gaan. De kans is dan groot dat de uitkomsten een vertekend beeld opleveren over het gemiddelde theaterbezoek. Door aan de resultaten van de alleenstaanden een relatief groter gewicht toe te kennen, kan deze vertekening worden verminderd.

Ook kan gebruik worden gemaakt van modellen om in te kunnen schatten of er vertekening is opgetreden en welke vertekening uiteindelijk overblijft. In het meeste onderzoek naar nonrespons worden hiervoor twee modellen gebruikt, te weten: het weerstandscontinuümmodel en het participantenklassenmodel (Lin & Schaef-

fer, 1995). Het eerstgenoemde model gaat er van uit dat hoe groter de inspanning van de onderzoeker is om respondenten mee te laten doen, des te hoger de respons is. Volgens dit model vormen de respondenten een ééndimensionele schaal waarin de uiteindelijke nonrespondenten qua resultaten lijken op de respondenten die de meeste inspanning van de onderzoeker vereisten om aan het onderzoek mee te doen. In het tweede model gaat men er vanuit dat mensen om verschillende redenen meedoen en dat verschillende typen respondenten met betrekking tot de resultaten lijken op verschillende typen nonrespondenten. Zo lijken weigeraars bijvoorbeeld op mensen die eerst nee zeggen en dan toch meedoen, en lijken niet-bereikten op moeilijk-bereikbaren. Als deze modellen opgaan, kan daarmee in de analyses rekening worden gehouden. Het werken met deze modellen kan helpen bij het schatten van de vertekening van nonrespons. Echter, om (één van) deze modellen te kunnen toepassen is informatie nodig over de respondenten én over de nonrespondenten op basis waarvan men hen kan onderscheiden (naar mate van weerstand of naar type respondent).

Helaas is ook deze schatting niet voldoende als de reden van de nonrespons te maken heeft met de inhoud van de survey. Als mensen in een onderzoek naar gevoelens van veiligheid niet meedoen omdat ze liever geen vreemden binnenlaten, en dus ook de enquêteur niet, dan is de kans groot dat de nonrespons tot vertekening leidt. Dit betekent dat de onderzoeker zich geconfronteerd ziet met een contradictie in terminus: namelijk dat het effect van nonrespons op de surveyresultaten pas kan worden ingeschat als er surveyinformatie van de nonrespondenten beschikbaar is. In het proefschrift van Stoop (2005) worden er twee mogelijke instrumenten genoemd om in een vervolgonderzoek alsnog surveyinformatie van nonrespondenten te verkrijgen:

- de centrale vraagprocedure (aan de nonrespondenten één of enkele vragen stellen die de kern van het onderzoek raken);
- een verkorte vragenlijst voorleggen aan een steekproef van de nonrespondenten.

In haar proefschrift bespreekt Stoop een dergelijk responsvervolgonderzoek naar de effecten van nonrespons uitgevoerd in het kader van het Aanvullend Voorzieningsgebruik Onderzoek (AVO). Om de resultaten van de respondenten en nonrespondenten aan dit onderzoek te vergelijken heeft Stoop een steekproef genomen van de meest hardnekkige nonrespondenten van het AVO en hen een verkorte vragenlijst voorgelegd die slechts door één lid van het gezin hoefde te worden ingevuld. De respons op dit vervolgonderzoek was 80%. Voor de vergelijking van de resultaten maakte Stoop gebruik van een indirecte manier (worst-case-scenario) waarbij óf alle nonrespondenten óf geen enkele nonrespondent in een bepaalde categorie resultaten (bijvoorbeeld de categorie 'helemaal geen culturele participatie' of de categorie 'culturele omnivoor') worden meegenomen, zodat de theoretisch waarschijnlijke boven- en ondergrens van de desbetreffende categorie bepaald kon worden. Deze grenzen werden vervolgens naast de resultaten van het vervolgonderzoek gelegd zodat uitspraken gedaan konden worden over de robuustheid van

de oorspronkelijke resultaten ofwel over de nonresponsbias van het onderzoek. Een meer directe manier van vergelijking van respondenten en nonrespondenten door middel van een verschilmaat als de effectgrootte  $D$  (Nijdam, 2003a, p. 90) is ook mogelijk maar dan moeten de gebruikte meetinstrumenten in de vragenlijst wel aan bepaalde eisen voldoen. In de meeste surveys worden Likertschalen gebruikt die gebaseerd zijn op de klassieke testtheorie (KTT). Een belangrijke assumptie van de Likertschaal is dat alle items in gelijke mate bijdragen aan de onderliggende latente variabele. De respondenten geven voor de verschillende items aan in welke mate ze akkoord (of niet akkoord) gaan met het item. Hoewel het hier om een ordinale schaal gaat, gaat men er in analyses vaak vanuit dat de Likertschaal van intervalmeetniveau is (zie ook de kritiek van Bond & Fox (2001) in bijlage III.2).

Een Raschmodel daarentegen is een wiskundig model gebaseerd op de Item Response Theorie (IRT) die onder andere de unidimensionaliteit van een bepaald construct specificeert. Daarnaast zijn de meeteenheden even groot (log odds of logits) en kunnen alle items op dezelfde schaal worden afgebeeld (zie ook bijlage III.2). Een voor deze vervolgstudie andere zeer belangrijke eigenschap van meting volgens het Raschmodel is dat de meetwaarden populatieonafhankelijk zijn (Bond & Fox, 2001; Curda, 1997; Ter Laak & De Goede, 2003). Populatieonafhankelijkheid impliceert dat iemands meetwaarde, verkregen met een concrete test (item), kan worden vastgesteld onafhankelijk van het moeilijkheidsniveau van de items in de test (Drenth & Sijsma, 1990). Het feit dat de meetwaarden populatieonafhankelijk zijn, maakt het eveneens mogelijk om de zogenoemde 'ability' van een persoon te bepalen ook al heeft deze persoon niet op alle items van een schaal geantwoord (Wolfe & Chiu, 1997). Als een schaal voldoet aan de eisen van een Raschmodel maakt het dus in principe niet uit welke items aan een persoon worden voorgelegd en kunnen persoonlijke vaardigheden direct met elkaar vergeleken worden (Wolfe, 2000).

In dit onderzoek zal de nonrespons van het crosssectionele evaluatieonderzoek onder de loep worden genomen. Bij dit evaluatieonderzoek zijn de resultaten gebaseerd op de geretourneerde vragenlijsten van 468 respondenten van de 2019 verstuurde vragenlijsten. Dit betekent een nonrespons van 77%. De vraag is of er gronden zijn aan te nemen dat in dit onderzoek sprake is van MCAR-nonrespons. Tijdens het evaluatieonderzoek is geen informatie verzameld over de weerstand van respondenten of over het type respondenten zodat het niet mogelijk is de door Stoop (2005) genoemde modellen (het weerstandscontinuümmodel en het participantenklassenmodel) te gebruiken. Om dit te kunnen doen, had vanaf het begin van het crosssectionele onderzoek een contactformulier moeten worden bijgehouden met onder andere informatie over het aantal keer dat een participant benaderd is (via mail, post of bij een nabelactie), over de reden van niet-deelname en over de mate van onbereikbaarheid. Dit is niet gebeurd.

Omdat, zoals eerder gezegd, de samenhang tussen oorzaakvariabelen van (on)bereidwilligheid complex is en de mogelijk bepalende variabele lastig te identificeren, zijn de respondenten en nonrespondenten voor representativiteit slechts

vergeleken op de algemene kenmerken sekse, leeftijd en opleiding (tabel 1). In deze tabel blijken de verschillen niet groot te zijn.

Tabel 1 Vergelijking respondenten en nonrespondenten naar sekse, leeftijd en opleiding

Variabele		Respondenten		Nonrespondenten		% verschil
		aantal	%	aantal	%	
Sekse	m	123	26	428	32	- 6
	v	345	74	1123	68	6
	<i>totaal</i>	<i>468</i>	<i>100</i>	<i>1651</i>	<i>100</i>	
Leeftijd	t/m 30	88	19	440	27	- 8
	31-40	171	37	627	37	0
	41-50	148	32	459	28	4
	51-	61	12	125	8	4
	<i>totaal</i>	<i>468</i>	<i>100</i>	<i>1651</i>	<i>100</i>	
Opleiding	1/2*	16	2	28	2	0
	3	31	6	111	7	-1
	4	52	11	174	10	1
	5	56	12	201	12	0
	6	68	15	209	13	2
	7	197	44	747	45	- 1
	8	48	10	181	11	- 1
	<i>totaal</i>	<i>468</i>	<i>100</i>	<i>1651</i>	<i>100</i>	

\* Code omschrijving voor in Nederland gevolg onderwijs

- 1 lager onderwijs, basisonderwijs
- 2 lager beroepsonderwijs (lts, leao, lavo, vglo, lhno, huishoudschool, lagere landbouw/tuinbouw/nautische school, 3 jaar handels-dag-avondschool, middenstandsdiploma, praktijkdiploma boekhouden)
- 3 algemeen voortgezet onderwijs (mavo, ivo, (m)ulo, middenschool, 3 jaar havo/mms/lyceum/hbs/vhmo/ atheneum/gymnasium)
- 4 middelbaar beroepsonderwijs (mts, voor '68 uts, meao, bestuursambt. ga-1, mhno, inas, verpleegkundige, kms, politieschool, middelbare detailhandelsschool, mba/spd-a, akte kleuter- of hoofdkleuterleidster, 5-jarige handelsavondschool)
- 5 hoger algemeen voortgezet onderwijs (havo, mms)
- 6 voorbereidend wetenschappelijk onderwijs (vwo, hbs, gymnasium, lyceum, atheneum)
- 7 hoger beroepsonderwijs (hts, voor '57 mts, heao, hoger bestuursambt. ga-1, spd-2, accountant adm. consulent, hogere landbouw/tuinbouw/zeevaartschool, hogere detailhandelsschool, nlo, mo-a (2<sup>e</sup>/3<sup>e</sup> graad), politie-academie, sociale/pedagogische academie, kma, hogere hotelschool, nwit, e.d.)
- 8 wetenschappelijk onderwijs (doctoraal opleiding, drs. ir, mr, arts, tandarts, apotheker, dr, nivra-accountant, mo-b (1<sup>e</sup> graad)
- 9 Onbekend

Het percentage vrouwen bij respondenten is licht hoger (6 %) dan bij de non-respondenten. De leeftijdscategorie tot en met 30 jaar is iets minder (8%) vertegenwoordigd bij de respondenten dan bij de nonrespondenten terwijl dit bij de leeftijdscategorieën vanaf 41 jaar andersom is. Voor wat betreft opleiding zijn er tussen de respondenten en nonrespondenten nauwelijks verschillen. Dit laatste is belangrijk omdat uit onderzoek van Tempelaar (2002) blijkt dat juist vooropleiding van invloed kan zijn op verschillen in leerresultaten.

#### *Vervolgonderzoek met centrale vraagprocedure of verkorte vragenlijst?*

In het crosssectionele onderzoek is gewerkt met een uitgebreide vragenlijst bestaande uit 164 vragen met betrekking tot attitude, motivatie, leerstrategieën en leeruitkomsten zoals functionele kennis van de statistiek. Gezien de diversiteit en hoeveelheid aan vragen ligt de centrale vraagprocedure niet voor de hand en is ervoor gekozen om een verkorte vragenlijst voor te leggen aan een aselechte steekproef van de nonrespondenten.

Volgens Stoop (2005) is er een grotere kans op vertekening als de nonrespons samenhangt met het onderwerp van het onderzoek. Bij het crosssectionele onderzoek zou de kans op vertekening groter kunnen zijn als de reden voor niet-deelname vooral te maken heeft met het feit dat men een hekel heeft aan statistiek. Als men vooral situationele redenen heeft voor niet-deelname (geen tijd, nog niet aan cursus begonnen, te lange vragenlijst) dan is de kans op vertekening kleiner. In de verkorte vragenlijst is daarom een extra vraag opgenomen over de reden van niet-deelname.

Volgens de 'rational choice theory of decision-making' (Groves & Couper, 1998) zetten individuen de kosten van deelname aan een survey af tegen de waargenomen voordelen. Om die reden wordt, om de respons te verhogen, soms gebruik gemaakt van kleine beloningen. Ook in de 'social exchange theory' (Stangor, 2004) gaat men ervan uit dat de beslissing om mee te doen aan een onderzoek mede wordt bepaald door de netto uitkomst van voor- en nadelen, maar deze voordelen hoeven niet per definitie materieel te zijn (Brehm, 1993). Niet-materiële voordelen kunnen legio zijn zoals de contacten met de onderzoeker, het feit dat men geïnteresseerd is in zijn of haar mening, de mogelijkheid om invloed te kunnen uitoefenen en het krijgen van feedback over de uitkomsten. Dit laatste zou ook van toepassing kunnen zijn op de respondenten in dit onderzoek. Bij het crosssectionele onderzoek gaf een aantal van hen namelijk aan graag geïnformeerd te worden over de (eigen) resultaten van het onderzoek. Het gaat hier om studenten die een bewuste keuze hebben gemaakt om (weer) te gaan studeren, om mensen die goed willen presteren en daarom ook nieuwsgierig zijn naar hun eigen prestaties. Wellicht wordt feedback over de eigen resultaten daarom ervaren als een beloning en wordt deelname aan het onderzoek op deze manier gestimuleerd. Om te onderzoeken of een dergelijke feedback bijdraagt aan de verhoging van de respons is aan de helft van de respondenten de toezegging gedaan dat de resultaten zullen worden teruggekoppeld. Bij de andere helft is deze toezegging weggelaten.

De onderzoeksvraag luidt:

*‘Leidt de lage respons in het crosssectionele evaluatieonderzoek tot vertekening van de resultaten?’*

Verwacht wordt dat vooral situationele factoren, in het bijzonder de zeer lange vragenlijst, als redenen worden opgegeven voor niet-deelname aan het crosssectionele onderzoek. Om die reden wordt verwacht dat er geen vertekening is opgetreden en dat de resultaten van de respondenten van het crosssectionele onderzoek en de respondenten van het responsonderzoek op de afhankelijke variabelen met elkaar overeenkomen. Daarnaast wordt verwacht dat de toezegging om feedback over de geleverde prestaties te geven, zal leiden tot een hogere respons in de desbetreffende groep. Voor een beschrijving van het determinantenmodel en de variabelen wordt verwezen naar paragraaf 5.2 van dit proefschrift. De hypothesen op basis van dit model luiden als volgt:

## Methode

### *Respondenten*

De onderzoeksgroep bestaat uit de nonrespondenten (1651) van het hoofdonderzoek. De groep is onderverdeeld in de competentiegerichte groep (641) en de vakgerichte groep (910). Uit elke groep is een aselechte steekproef getrokken. Van de in totaal 742 verstuurdde vragenlijsten zijn er 210 (28%) geretourneerd. Hiervan hebben 49 respondenten alleen de motievenvraag (reden niet-deelname crosssectionele onderzoek) ingevuld. Tabel 2 geeft de verdeling over de vier onderscheiden groepen weer.

Tabel 2 Steekproef en responsaantallen

	<i>totale groep nonrespondenten</i>	<i>aselecte steekproef vervolgonderzoek</i>	<i>respons</i>	<i>%</i>	<i>alleen motievenvraag</i>	<i>respons</i>
COMP	641	286	95	33%	19	76
VAK	910	456	115	25%	30	85
Totaal	1651	742	210	28%	49	161

### *Procedure*

De 742 geselecteerde studenten uit de populatie nonrespondenten ( $N = 1651$ ) hebben per e-mail (563) of per post (179) een verkorte versie van de oorspronkelijke vragenlijst toegestuurd gekregen met het verzoek de vragen te beantwoorden en zo spoedig mogelijk te retourneren. Bij de postversie is een voorgefrankeerde en geadresseerde antwoordenvolp bijgevoegd. Na twee weken is aan de nonrespondenten nog een herinnering gestuurd.



### *Meetinstrument*

Het onderzoeksinstrument bestaat uit een korte vragenlijst met 13 vragen die afkomstig zijn uit de crosssectionele vragenlijst (bijlage II.2). Van de 13 vragen worden 11 vragen met betrekking tot statistiek en onderzoekstaken beantwoord op een vijfpuntsschaal. De antwoordcategorieën variëren van 1 (geheel oneens) tot 5 (geheel mee eens). Voor de vragenlijst van dit nonresponsvervolgonderzoek zie Appendix A. Met de vragen 1 tot en met 11 worden verschillende concepten gemeten zoals Intrinsieke waarde (vraag 1), Testangst (vraag 2), Zelfeffectiviteit (vragen 3, 4 en 5), Zelfregulatie (vragen 6 en 7), Kritisch denken (vragen 8 en 9), Affect (vraag 10) en Cognitieve competentie (vraag 11). De keuze voor de verschillende vragen is gemaakt op basis van de Raschanalyses van de schalen van de crosssectionele vragenlijst. Bij de keuze van de items was de positie op de schaal bepalend (in verband met optredende meetfouten bij items bij voorkeur rond het midden) en in het geval van meer items eveneens de positie van de items ten opzichte van elkaar (voldoende onderscheidend van elkaar).

Vraag 12 vraagt naar de reden van niet-deelname aan het crosssectionele onderzoek door middel van vier vaste antwoordmogelijkheden en één antwoordmogelijkheid waarbij men zelf een reden kan aangeven.

Vraag 13 is een open kennisvraag waarin gevraagd wordt de vuistregel voor de standaardafwijking te beschrijven in termen van gemiddelde, standaardafwijking en verdeling en in de beschrijving zoveel mogelijk argumenten te noemen (zie ook item 140 in de crosssectionele vragenlijst). De antwoorden en argumenten op deze vraag zijn geëvalueerd en geclassificeerd op basis van de Structure of Observed Learning Outcomes (SOLO) taxonomie van Biggs en Collis (1982). Binnen de SOLO-taxonomie worden vijf niveaus van complexiteit onderscheiden. De SOLO-scores zijn beoordeeld en geclassificeerd door twee onafhankelijke beoordelaars aan de hand van een antwoordmodel waarbij verschillende mogelijke antwoorden zijn ingedeeld volgens de SOLO-hiërarchie (zie paragraaf 5.2 van dit proefschrift). Van de vragenlijst voor de nonrespondenten zijn twee versies gemaakt: één met toezegging voor feedback en één zonder deze toezegging.

### *Analyse*

De antwoorden op de open kennisvraag zijn door twee beoordelaars onafhankelijk van elkaar beoordeeld. De eerste vergelijking van de coderingen leverde voor Cohen's Kappa een waarde op van .64. Een kritische tweede beoordeling verhoogde de interbeoordelaarsbetrouwbaarheid naar een Kappa-waarde van .86. De items 9 (Kritisch denken) en 10 (Affect) zijn gehercodeerd.

In het programma WINSTEPS (Linacre, 2005) is vervolgens de moeilijkheid van de items met behulp van Raschanalyses gefixeerd ('anchoring') op de verkregen Raschmeetwaarden (difficulty measures) uit het crosssectionele onderzoek (Wolfe, 2000). Dit levert Raschmeetwaarden per persoon op (person measures) voor de items en Raschmeetwaarden per persoon en item (person- en itemmeasures) voor itemclusters. Vervolgens zijn de desbetreffende items uit het crosssectionele onderzoek geïsoleerd en op dezelfde manier geijkt op de eerdergenoemde itemmoeilijk-



heid. De scores op de items 3, 4 en 5 zijn in één analyse geïkt en samengevoegd tot één itemcluster Zelfeffectiviteit, de scores op de items 6 en 7 zijn op deze wijze samengevoegd tot één itemcluster Zelfregulatie en de scores op de items 8 en 9 tot één itemcluster Kritisch denken. Na deze analyses is het mogelijk de personen (person abilities) uit het responsvervolgonderzoek op dezelfde schaal te vergelijken met de personen (person abilities) uit het crosssectionele onderzoek. Ter illustratie is dit voor de itemclusters Zelfeffectiviteit, Zelfregulatie en Kritisch denken grafisch weergegeven in de map van personen en items (Appendix B).

De vergelijking tussen het crosssectionele onderzoek en het responsvervolgonderzoek vindt plaats op itemniveau omdat voor een vergelijking op schaalniveau sprake moet zijn van een itembank en een minimaal aantal items per schaal (Wright & Bell, 1984). Als maat voor het verschil wordt de effectgrootte  $D$  gehanteerd, zijnde het verschil tussen de gemiddelde waarden gedeeld door de standaardafwijking. De interpretatie van  $D$  is als volgt:  $D < 0.4$ : gering verschil;  $0.4 \leq D < 0.8$ : middelmatig verschil;  $D \geq 0.8$ : groot verschil (Nijdam, 2003a, p. 91). Met de chikwadraattoets is beoordeeld of de toezegging om feedback te geven, samenhangt met de uiteindelijke respons in de verschillende condities.

### Resultaten

De 49 vragenlijsten waarop alleen de reden voor het niet-deelnemen aan het crosssectionele onderzoek is aangegeven, zijn bij de toetsing van de hypothesen 1 tot en met 3 buiten beschouwing gelaten. Voor de toetsing van de hypothese 1 tot en met 3 bleven om die reden 161 (22%) respondenten over. De respondenten en de nonrespondenten zijn vergeleken op sekse, leeftijd en opleiding (tabel 3).

Tabel 3 Vergelijking respondenten en nonrespondenten uit de nonresponsgroep naar sekse, leeftijd en opleiding

Variabele		Respondenten		Nonrespondenten		% verschil
		aantal	%	aantal	%	
Sekse	m	60	29	161	30	- 1
	v	150	71	371	70	1
	<i>totaal</i>	210	100	532	100	
Leeftijd	t/m 30	40	19	142	27	- 8
	31-40	78	38	196	37	1
	41-50	73	34	147	27	7
	51-	19	9	47	9	0
	<i>totaal</i>	210	100	532	100	
Opleiding	1/2*	2	0	11	2	- 2
	3	13	6	39	7	- 1
	4	26	12	63	12	0
	5	24	11	62	12	- 1
	6	21	10	69	13	- 3
	7	98	47	229	43	4
	8	26	12	59	11	1
	<i>totaal</i>	210	100	532	100	

\* Code omschrijving voor in Nederland gevolg onderwijs, zie tabel 1.

Er zijn geen noemenswaardige verschillen gevonden voor wat betreft sekse. Voor wat betreft leeftijd is een zelfde verschil te zien als bij de respondenten en de nonrespondenten uit het crosssectionele onderzoek. De leeftijdscategorie tot en met 30 jaar is minder vertegenwoordigd bij de respondenten dan bij de nonrespondenten; bij de leeftijdscategorie 41 tot en met 50 jaar is dit andersom. Voor wat betreft opleiding zijn er kleine verschillen te zien in de categorieën 6 (vwo) en 7 (hbo).

Wat betreft de verschillen tussen de groepen op de afhankelijke variabelen, wordt verwezen naar de tabellen 4 en 5. Tabel 4 bevat de gemiddelden en standaarddeviaties van de scores van de respondenten van het crosssectionele onderzoek en van de 'nonrespondenten' uit het responsvervolgonderzoek binnen COMP en VAK op de geselecteerde item(cluster)s. Tabel 5 geeft de verschillen tussen de diverse te onderscheiden groepen weer, uitgedrukt in de effectmaat Cohen's *D*. Tabel 4 laat zien dat de richting van de verschillen in gemiddelde waarden van COMP en VAK in het crosssectionele als het nonresponsonderzoek op twee na alle overeenkomen, behalve bij Intrinsieke waarde als Testangst. Hoewel de gemiddelde waarden op de item(cluster)s volgens tabel 4 voor COMP gunstiger zijn dan voor VAK – hetgeen volgens de verwachting is - blijken de verschillen in effectgrootten gering te zijn, zowel in het responsvervolgonderzoek als in het crosssectionele onderzoek. Ook de verschillen tussen beide COMP-groepen en tussen de beide VAK-groepen zijn op een enkele variabele na gering. Er zijn derhalve geen aanwijzingen dat de groepen systematisch verschillen op de afhankelijke variabelen.

Tabel 4 Gemiddelden en standaarddeviaties van de 'person measures' op de item(cluster)s in COMP en VAK in Rasch logits

Groepen	COMP Crosssectioneel onderzoek n = 128		COMP Responsonderzoek n = 76		COMP Crosssectioneel onderzoek n = 340		COMP Responsonderzoek n = 85	
	M	SD	M	SD	M	SD	M	SD
Item(cluster)s								
Intrinsieke waarde	.869	.995	.764	1.154	.514	1.067	.962	.951
Testangst	-1.196	.927	-.371	1.104	-.927	1.153	-.476	.968
Zelfeffectiviteit	4.692	2.958	3.678	2.770	3.455	3.919	3.211	2.550
Zelfregulatie	2.417	2.974	2.690	3.013	1.643	2.888	1.809	2.449
Kritisch denken	2.549	2.691	3.919	2.717	1.708	2.652	3.792	3.319
Affect	.546	1.321	.255	1.413	-.026	1.399	.146	1.344
Cognitieve comp	.850	.764	.948	.853	.468	1.187	.799	.923
Statistiekennis	-.067	1.297	-.437	1.299	-.439	1.281	-.766	1.352

Tabel 5 VerschilmatenCohen's *D* voor de verschillende item(cluster)s tussen COMP en VAK

	<i>COMP - VAK Responsonderzoek</i>	<i>COMP - VAK Crosssectioneel onderzoek</i>	<i>COMP - COMP Responsonderzoek</i>	<i>VAK - VAK Crosssectioneel onderzoek</i>
<i>Item(cluster)s</i>	<i>D</i>	<i>D</i>	<i>D</i>	<i>D</i>
Intrinsieke waarde	.18	.34	.11	.42
Testangst	.11	.16	.28	.42
Zelfeffectiviteit	.16	.35	.36	.07
Zelfregulatie	.34	.28	.09	.06
Kritisch denken	.04	.32	.51	.69
Affect	.08	.42	.22	.08
Cognitieve comp	.16	.38	.12	.31
Statistiekennis	.25	.29	.28	.25

De reden voor niet-deelname aan het crosssectionele onderzoek blijkt ook hier voornamelijk situationeel (vragenlijst te lang, geen tijd, gestopt met de studie) en nauwelijks iets te maken met het onderwerp van het onderzoek (hekel aan statistiek). Van de 210 respondenten geven 195 respondenten een situationele reden op (93%), 3 respondenten (1%) hebben geen reden opgegeven en 12 respondenten (6%) hebben als reden opgegeven een hekel aan statistiek te hebben. Ook ten aanzien van de feedbackconditie blijken er geen verschillen te zijn.

## Discussie

De onderzoeksvraag luidde: 'Leidt de lage respons in het evaluatieonderzoek tot vertekening van de resultaten?' Uit de verschillende vergelijkingen tussen COMP en VAK wat betreft sekse, leeftijd, vooropleiding en de patronen in de afhankelijke variabelen zijn geen aanwijzingen gevonden dat de lage respons tot vertekening ('nonresponsbias') heeft geleid. De reden voor niet-deelname bleek vooral situationeel. Een punt van discussie is wel het tijdstip van de meting van de afhankelijke variabelen in het nonresponsonderzoek. Deze vond plaats na het crosssectionele onderzoek van COMP. Het was voor de studenten VAK dus ruim een jaar langer geleden dan voor de studenten uit COMP dat men de desbetreffende modules bestudeerd had. Dit kan (mede) van invloed zijn geweest op de scores en een verklaring zijn voor de gevonden verschillen. Een ander punt van discussie is dat ook met de zeer verkorte vragenlijst de respons laag was (23%).

## Aanbevelingen

Met Stoop (2005) kan men van mening zijn dat er bij surveyonderzoek altijd rekening moet worden gehouden met een mogelijke vertekening door nonrespons. De hoogte van de initiële respons is hierbij niet van belang omdat de omvang van nonrespons geen goede maat is voor de mate van vertekening. Het vervolgonderzoek met een verkorte vragenlijst waarbij respondenten en nonrespondenten direct met

elkaar vergeleken worden, kan een goed instrument zijn om deze vertekening in te schatten. Dit betekent dan wel dat het vervolgonderzoek al vanaf het begin een plaats in de onderzoeksopzet moet krijgen.

Zo moet duidelijk zijn of de schalen van de gebruikte vragenlijsten voldoen aan de eisen die het Raschmodel er aan stelt, of er sprake is van een itembank en ook hoeveel items er minimaal nodig zijn om een bepaald construct te meten zodat een doelmatige, verkorte versie van de vragenlijst geconstrueerd kan worden.

Ook kan tijd een belangrijke factor zijn. Als er veel tijd zit tussen het hoofdonderzoek en het vervolgonderzoek dan zullen de resultaten in bepaalde gevallen wellicht eerder vertekening opleveren. Dit kan ook in het responsonderzoek een rol gespeeld hebben. Responsvervolgonderzoek zou dan - zeker als het gaat om het meten van leerresultaten - direct moeten plaatsvinden na het hoofdonderzoek zodat deze invloed geminimaliseerd wordt.

### Literatuur

- Babbie, E. (2007). *The Practice of Social Research* (11<sup>th</sup> ed.) Belmont, CA: Wadsworth.
- Biggs, J.B., & Collis, K.F. (1982). *Evaluating the Quality of Learning: the SOLO taxonomy*. New York: Academic Press.
- Bond, T.G., & Fox, C.M. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. Londen: Lawrence Erlbaum Associates, Publishers.
- Brehm, J. (1993). *The phantom respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.
- Curda, L.K. (1997). *Validation of a measure of teachers' efficacy and outcome expectations in the content domains of reading and mathematics*. Dissertation. Norman: University of Oklahoma.
- Drenth, P.J.D., & Sijtsma, K. (1990). *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen*. Houten/Diegem: Bohn Stafleu Van Loghum.
- Ellemers, J.E. (2004). Wat we nog niet weten. Tekortkomingen en blinde vlekken bij de analyse van de Nederlandse maatschappij. *Sociologische gids* 51(4), 482-487.
- Groves, R.M., & Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Lin, I., & Schaeffer, N.C. (1995). Using survey participants to estimate the impacts of non-participation. *Public Opinion Quarterly*, 59, 236-258.
- Stangor, C. (2004). *Social groups in action and interaction*. New York: Psychology Press.
- Stoop, I.A.L. (2005). *The hunt for the last respondent: nonrespons in sample surveys*. Den Haag: Sociaal Cultureel Planbureau (SCP).
- Tempelaar, D. (2002). *Modelling Students' Learning Of Introductory Statistics*.

Retrieved

January 12, 2006 from [http://www.stat.auckland.ac.nz/~iase/publications/1/6f1\\_temp.pdf](http://www.stat.auckland.ac.nz/~iase/publications/1/6f1_temp.pdf).

Ter Laak, J.J.F., & De Goede, M.P.M. (2003). *Psychologische diagnostiek. Inhoudelijke en methodologische grondslagen*. Lisse: Swets & Zeitlinger.

Viswesvaran, C., Barrick, M.R., & Ones, D.S. (1993). How definitive are conclusions based on survey data: estimating robustness to nonresponse. *Personnel Psychology*, 46, 551-567.

Wolfe, E.W. & Chiu, C.W.T. (1997). *Measuring change over time with a Rasch rating scale model*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, USA.

Wolfe, E.W. (2000). Equation and Item Banking with the Rasch Model. *Journal of Applied Measurement*, 1(4), 409-434.

Wright, B.D., & Bell, S.R. (1984). Item banks: What, why, how. *Journal of Educational Measurement*, 2, 331-345.

## Appendix A Vragenlijst

Geachte student,

In uw opleiding komt u cursussen tegen die betrekking hebben op het doen van onderzoek en de statistische verwerking van gegevens. Door middel van de vragen in deze vragenlijst wil-len we uw mening peilen over deze cursussen. Het zijn in totaal 13 vragen waarvan 11 vragen waarop u uw antwoord kunt geven door in één vakje een X te zetten, één motievenvraag en één kennisvraag. In totaal zal het niet meer dan 5 minuten van uw tijd vragen. Wij stellen uw medewerking zeer op prijs en daarom zullen we u na afloop van het onderzoek informeren over de resultaten van het onderzoek in het algemeen en uw eigen resultaten in het bijzonder (*alleen voor feedbackconditie*). Bij voorbaat onze dank voor uw deelname aan dit onderzoek. In verband met de verwerking van de gegevens verzoeken wij u eerst de volgende gegevens op te geven:

Studentnummer (9 cijfers) .....

Postcode en huisnummer .....

	Svp bij elke vraag in één vakje een X zetten	1 geheel oneens	2 oneens	3 oneens noch eens	4 mee eens	5 geheel mee eens
1	Als ik onderzoek doe, ga ik helemaal op in het onderzoeks-proces (flow)					
2	Als ik mijn eindopdracht maak, denk ik eraan hoe slecht ik het doe (ta).					
3	Ik ben er zeker van dat ik de onderzoekstaken heel goed zal kunnen uitvoeren (se)					
4	Vergeleken met andere studenten die onderzoek doen en daarbij statistiek toepassen, verwacht ik het beter te zullen doen. (se)					
5	Ik denk dat ik in staat ben om de theorie uit de Methoden & Technieken modules te begrijpen en daardoor onderzoek uit te kunnen voeren (se).					
6	Bij het uitvoeren van nieuwe onderzoekstaken maak ik gebruik van wat ik heb geleerd in andere opdrachten (sr).					
7	Als een onderzoekstaak niet verplicht is, dan maak ik hem toch (sr).					
8	Ik focus altijd op de kwaliteit van de meetinstrumenten die men in een onderzoek gebruikt en ga na hoe betrouwbaar en valide die meetinstrumenten zijn (ct)					
9	Als ik onderzoeksverslagen lees kijk ik nooit naar de beschrijving van de onderzoeksgroep (ct)					
10	Statistiek tentamens frustreren mij (affect)					
11	Ik begrijp statistische vergelijkingen (cc)					

## 12 Motievenvraag

In een eerder stadium heeft u reeds een vragenlijst over de statistiek- en onderzoeksmodules van de OU ontvangen. Aan dat evaluatieonderzoek heeft u niet meegedaan. Kunt u aangeven welke van de onderstaande genoemde redenen het meest op u van toepassing was? S.v.p. slechts één antwoord aankruisen (één kruisje (X) in het blokje zetten).

- ☐ ik had geen tijd
- ☐ ik vond de vragenlijst te lang
- ☐ ik heb een hekel aan statistiek
- ☐ ik was nog niet begonnen aan de cursus
- ☐ anders, namelijk .....

## 13 Open vraag

Wat weet u op dit moment nog van de volgende statistische begrippen?

Welke relatie bestaat er volgens u tussen de statistische termen *gemiddelde*, *standaardafwijking* (ook wel standaarddeviatie genoemd) en *normaalverdeling*? *normaalverdeling* bij de beschrijving van een bekende vuistregel over de standaardafwijking?

Beschrijf eventuele relaties in eigen woorden. Geef zoveel mogelijk argumenten.

.....

.....

.....

.....

.....

.....

.....

.....

.....

Einde vragenlijst

## Appendix B Gecombineerde map van personen en items

### Zelfeffectiviteit

	<more>		<rare>		
14	#	+			
13	.	+			
12		+			
11		+			
10	#	T	+		
9		+			
	#	T			
8	####	+	T		
7		S	+		
	###				
6	#####	S	+	T	VAR4 selfefficacy 2
			S		
5	#####	+			
			S	VAR4	selfefficacy 2
4	#####	M	+		
	#####	M		M	VAR3 selfefficacy 1
3		+			
	#####		M	VAR3	selfefficacy 1
2	#####	+			
	#####				
1	###	S	+	S	
		S		S	VAR5 selfefficacy 3
0	###	+			
	##				
-1	#	T		T	
-2		T			
	#	+			
-3					
-4		+			
-5	#	+			
-6		+			
-7		+			
	<less>		<frequ>		

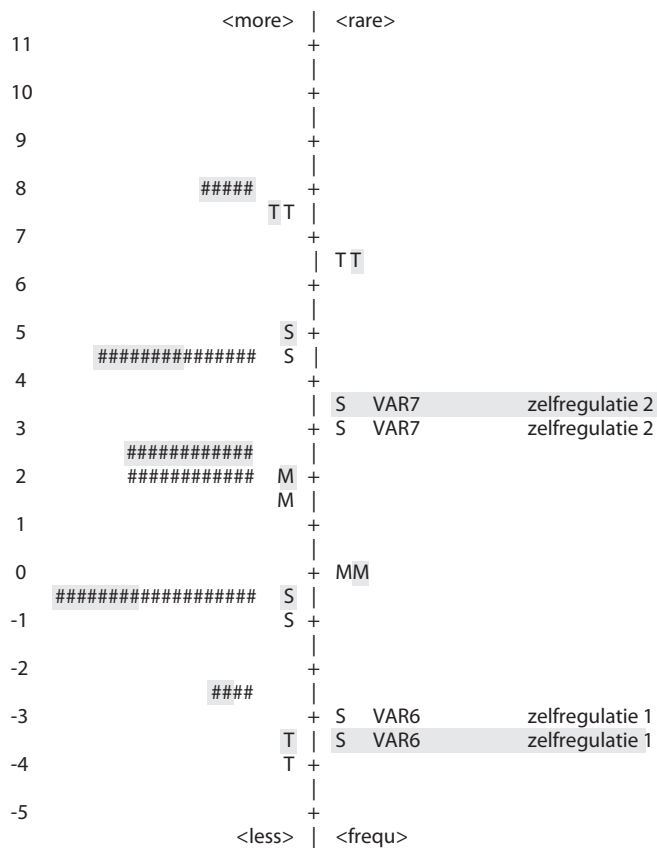
EACH '#' IS 8; EACH '#' IS 5.

zwart = crosssectionele onderzoek

in grijs = responsonderzoek



## Zelfregulatie

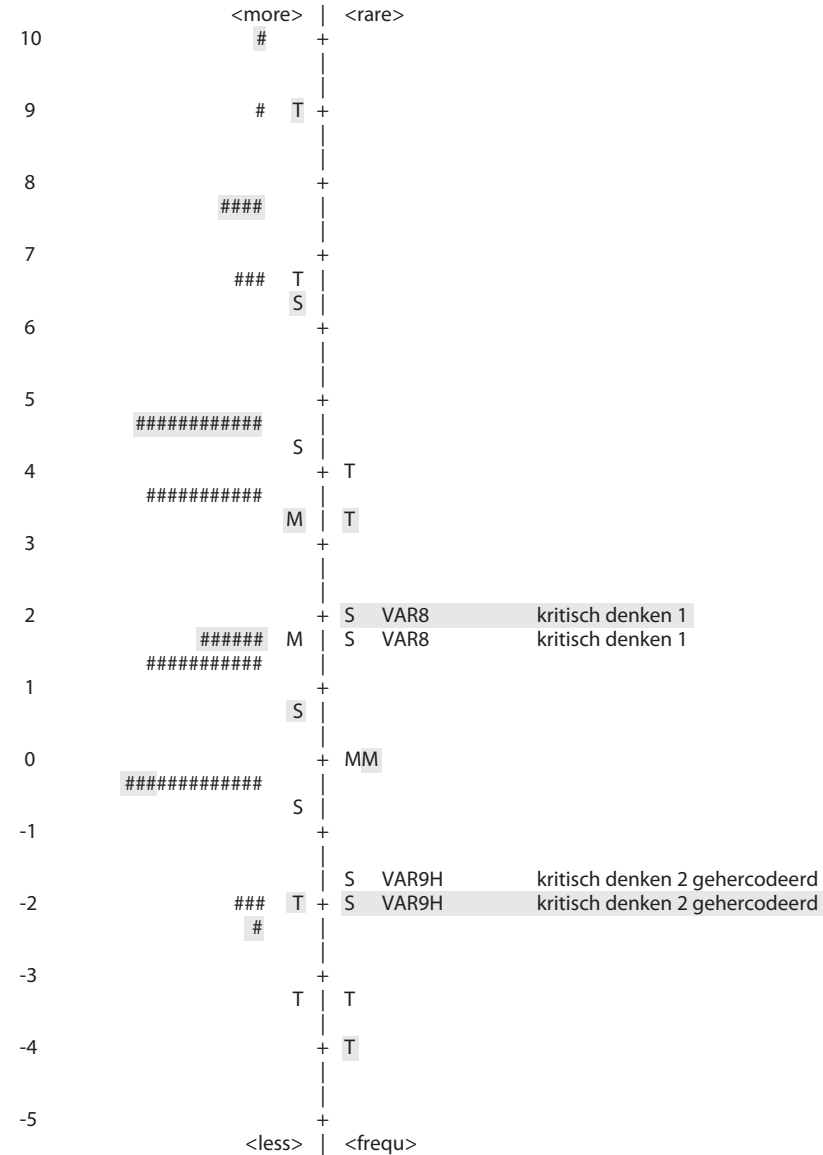


EACH '#' IS 12; EACH '#' IS 5.

zwart = crossectionele onderzoek

in grijs = responsonderzoek

Kritisch denken



EACH '#' IS 11; EACH '#' IS 6 .

zwart = crosssectionele onderzoek  
in grijs = responsonderzoek



## **Voorstudie Raschanalyses van de oorspronkelijke Likertschalen**

## Abstract

Deze voorstudie beschrijft de redenen waarom en de wijze waarop de Likertschalen die zijn gebruikt in het proefschriftonderzoek volgens het Raschmodel zijn geanalyseerd. Eerst wordt uiteengezet welke bezwaren er bestaan tegen het toepassen van de klassieke testtheorie bij Likertschalen. Vervolgens wordt in de methodenparagraaf beschreven welke criteria bij Rasch-analyses worden gehanteerd om te kunnen spreken van een ééndimensionale meetschaal. In de resultatenparagraaf wordt bij elke schaal nagegaan in hoeverre deze voldoet aan de Raschmodelcriteria. Bij niet goed passen zijn er voorstellen gedaan en uitgevoerd om alsnog een passende schaal te verkrijgen.

## Introductie

De schalen die in dit proefschrift voor de vaststelling van attitude, motivatie en leerstrategieën in het determinantenmodel zoals beschreven in hoofdstuk 2 zijn gebruikt, zijn alle van het type Likertschalen met vijf antwoordcategorieën. Binnen veel psychologisch onderzoek is het gebruikelijk deze volgens de klassieke testtheorie (KTT) te toetsen op ééndimensionaliteit met confirmatorische factoranalyse (CFA), vervolgens te beoordelen op interne consistentie met Cronbach's alpha en ten slotte de somscore te beschouwen als de mate van aanwezigheid van de door de schaal bedoelde psychologische eigenschap. Tegen deze handelswijze zijn een aantal bezwaren te noemen (zie onder anderen Bond & Fox, 2001; Curda, 1997; Ter Laak & De Goede, 2003). Ten eerste valt uit een bepaalde somscore geen uniek of geordend antwoordpatroon te herleiden. Een respondent kan daardoor op veel manieren een bepaalde somscore halen. Aan de hand van een praktijkvoorbeeld uit het huidige onderzoek wordt dit probleem verduidelijkt.

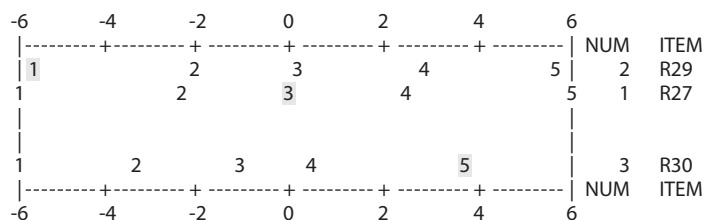
Bij het construeren van een meetinstrument, bijvoorbeeld een schaal voor testangst, is de aanname dat hoge totaalscores een hoge mate van testangst reflecteren en lage totaalscores een lage mate van testangst. De items worden gescoord op een vijfpuntschaal die een bereik heeft van 1) geheel mee oneens 2) mee oneens 3) noch oneens, noch eens 4) mee eens, tot 5) geheel mee eens. Zie bijvoorbeeld de volgende drie items van de schaal voor Testangst:

Item 29: 'Ik ben zo gespannen tijdens het maken van mijn eindopdracht dat ik de dingen die ik geleerd heb niet meer op de juiste wijze toe kan passen.'

Item 27: 'Is ik mijn eindopdracht maak denk ik eraan hoe slecht ik het doe.'

Item 30: 'Ik maak me zorgen over de eindopdrachten.'

In figuur 1 zijn twee fictieve personen A en B weergegeven, die verschillende scores hebben op deze drie items. Persoon A is vertegenwoordigd met de vet gedrukte scores en persoon B met de grijs gemarkeerde scores. Van onder naar boven geeft een Rasch-diagnose de moeilijkheidshiërarchie van de items weer, wat in dit voorbeeld betekent dat item 30 het gemakkelijkste item is (dit item is door de meeste respondenten met 'mee eens' of 'geheel mee eens' beantwoord), veel gemakkelijker te onderschrijven dan item 27, dat moeilijker te beamen is en op zijn beurt toch weer minder moeilijk is dan item 29, dat het moeilijkst te onderschrijven item van deze kleine reeks is.



Figuur 1 Fictieve persoon A en B in een Raschmodeldiagnose

Aan persoon A, met de scores 2, 3 en 4, wordt een totaalscore van 9 toegekend, die vervolgens wordt toegepast in vervolganalyses. Aan persoon B, met de scores 5, 3 en 1, wordt eveneens een totaalscore van 9 toegekend, hoewel deze persoon een volstrekt ander antwoordpatroon laat zien.

Ten tweede is een somscore alleen gerechtvaardigd als elk item uitwisselbaar is en in dezelfde mate bijdraagt aan de bepaling ervan. In de praktijk is van deze aanname bijna nooit sprake, getuige de uiteenlopende item-totaalcorrelaties (Curda, 1997) en de ongelijke bijdragen van items op de factorladingen (Byrne, 2001). Items verschillen in moeilijkheidsgraad, gematigdheid, extreemheid of populariteit waardoor het voor respondenten niet vanzelfsprekend is ze alle gelijkelijk te behandelen. Op respondenten wordt daardoor niet altijd een gelijk beroep gedaan op het niveau van de psychologische eigenschap.

Het voorbeeld in figuur 1 toont dat de aanname van 'gelijke bijdragen' niet opgaat. Het is veel eenvoudiger om het veel meer algemeen geformuleerde item 30 te beamen dan de meer specifieke items 27 en 29, en het is iets minder moeilijk om te onderschrijven dat men aan zichzelf twijfelt tijdens het uitvoeren van de eindopdracht (item 27) dan om te beamen dat men helemaal in de stress schiet (item 29). Deze moeilijkheidsvolgorde wordt weerspiegeld in het antwoordpatroon van persoon B. Daarentegen valt op te merken dat persoon A het gemakkelijkste item nauwelijks onderschrijft, het één na moeilijkste item al wat meer onderschrijft en juist het moeilijkste item beaamt. De optelprocedure van de Likertschaal leidt tot de conclusie dat er geen verschil in testangst is tussen persoon A en B. De Raschprocedure, die onder andere onderscheid maakt in moeilijkheidsgraad van items, laat zien dat persoon A duidelijk meer testangst heeft dan persoon B, omdat juist zijn scores op de twee moeilijkste items het hoogst zijn.

Terecht argumenteren Bond & Fox (2001) dat de data zodanig moeten worden geanalyseerd dat de totaalscore op een construct de bijdrage van elk item aan het construct reflecteert.

De aanname van 'elk item levert een gelijke bijdrage' als rechtvaardiging voor het sommeren van de scores veronderstelt tevens meting op intervalniveau. Hierbij wordt ervan uitgegaan dat de afstanden tussen alle antwoordcategorieën gelijk zijn, zowel binnen als tussen de items. Er kunnen echter grote psychologische verschillen bestaan

tussen het onderschrijven van 'geheel oneens' tot 'geheel mee eens'. Dit gebrek aan lineariteit doet zich niet alleen voor binnen items maar ook tussen items.

Bijvoorbeeld, in figuur 1 is te zien dat de psychologische afstand tussen 'noch oneens/noch eens' en 'mee eens' in item 30 aanmerkelijk kleiner is dan de psychologische afstand tussen diezelfde categorieën in item 27 en item 29. Binnen item 30 zelf is het duidelijk dat de afstand tussen 'mee eens' en 'geheel mee eens' aanmerkelijk groter is dan de afstand tussen 'noch oneens/noch eens' en 'mee eens'. De grotere meetstap tussen 'mee eens' en 'geheel mee eens' is in ons voorbeeld evident in alle items van de schaal. Het voorbeeld maakt duidelijk dat de meetstappen tussen de categorieën sterk verschillen, zowel binnen items als tussen items en dat een Likertschaal bijgevolg niet meer is dan een ordinale schaal, waarvan de scores eerst getransformeerd moeten worden naar een intervalniveau met gelijke meetstappen, voordat ze opgeteld mogen worden (Bond & Fox, 2001).

Ten derde bestaat er bij Likertschaalprocedures het bezwaar dat prestaties van personen niet los van de gebruikte tests of schalen vergeleken worden. Indien twee tests worden gebruikt die dezelfde psychologische eigenschap meten maar niet gelijk zijn in moeilijkheidsgraad, dan is een vergelijking binnen en tussen de personen die de twee tests hebben ingevuld hebben, niet mogelijk. Een hiermee samenhangend bezwaar is dat de waarschijnlijkheid van 'succes' op een bepaald item gegeven een bepaalde persoon met een bepaalde bekwaamheid binnen de KTT niet kan worden bepaald. De KTT is gebaseerd op de assumptie dat de variantie van de fouten dezelfde is voor alle personen ondanks het feit dat het niet ongewoon is dat de ene persoon zich wat consistentier gedraagt dan de andere (Curda 1997).

Ten vierde bestaat er het bezwaar dat de intrinsieke betekenis van een somscore binnen de KTT alleen kan worden ontleend aan een norm die gebaseerd is op een steekproef die representatief is voor een bepaalde doelgroep. Deze norm kan echter voor elke doelgroep anders uitvallen. De testbetrouwbaarheid is direct gerelateerd aan de variabiliteit van de testcores (Curda, 1997). Met andere woorden: als de ware score van een bepaalde steekproef verandert ten opzichte van een eerdere steekproef, dan zal de betrouwbaarheidscoëfficiënt eveneens veranderen. Daarom zijn items en testparameters die gebaseerd zijn op de KTT alleen bruikbaar voor populaties die vergelijkbaar zijn met de steekproef waarop de parameters zijn gebaseerd (Curda, 1997). Hiermee vervalt de mogelijkheid van het equivaleren van items (Curda, 1997). Voor dit proefschrift is deze equivalencemogelijkheid op tweeërlei wijze van belang. Ten eerste bij de beschouwing van de nonrespons van dit onderzoek (zie hiervoor hoofdstuk 5 van dit proefschrift en bijlage III.1) en ten tweede bij mogelijke vervolgstudies waarin uitgegaan moet worden van vragenlijsten die een (beperkt) aantal items gelijk hebben (zoals onder andere het geval is bij de longitudinale evaluatie in hoofdstuk 6 van dit proefschrift).

Modellen die gebaseerd zijn op de Item Response Theorie (IRT) komen aan bovengenoemde bezwaren tegemoet (Bond & Fox, 2001; Drenth & Sijtsma, 1990; Ter Laak & De Goede, 2003). Een van die modellen betreft het éénparametermodel van Rasch.

Met het Raschmodel kan worden gecontroleerd in hoeverre de scores van een groep personen op een verzameling van items passend kunnen worden beschreven. Indien dit laatste het geval is, dan volgt daaruit dat theoretische afleidbare eigenschappen van de meting ook in concrete, praktische gevallen gelden (Drenth & Sijsma, 1990, p. 131). Raschmodellen genereren metingen van personen en items op intervalniveau (meeteenheden zijn de zogeheten logits, of natuurlijke logaritmes van de odds) en impliceren additiviteit van de scores in het model. Een andere gunstige eigenschap van het model is dat iemands vaardigheidsniveau kan worden vastgesteld onafhankelijk van het moeilijkheidsniveau van de items in de test (of, indien er sprake is van een psychologische eigenschap, onafhankelijk van de mate van extreemheid, gematigdheid of populariteit van de items op een schaal). De mogelijkheid van *populatieonafhankelijk meten* houdt in dat het voor de vaststelling van de schaalwaarde van een persoon niet uitmaakt of de test moeilijk (of de schaal extreem) of gemakkelijk (of gematigd) is: de meting van een persoon op de schaal is vrij van het moeilijkheidsniveau van de test en iemands vaardigheidsniveau is in alle gevallen dezelfde. Uiteraard dienen de schalen wel te voldoen aan de strenge eisen die door het Raschmodel aan testgegevens worden opgelegd (zie hierna).

## Methode

### *Respondenten*

Studenten van de opleiding Psychologie van de Open Universiteit Nederland die zich in de studiejaar 2003-2004 en 2004-2005 hebben ingeschreven voor een cursus statistiek respectievelijk het onderzoekspracticum kwantitatieve data-analyse, zijn verzocht deel te nemen aan de evaluatie van deze onderwijsmodulen. Van de studenten die een statistiekcursus hebben gevolgd, hebben 340 studenten de vragenlijst geretourneerd en van de studenten die het onderzoekspracticum hebben gevolgd 128 studenten, in totaal 468 studenten.

### *Meetinstrumenten*

Studenten hebben een vragenlijst voorgelegd gekregen met daarin opgenomen een aantal vijfpunts- Likertschalen en een statistiekkennistest. De Likertschalen betreffen de volgende psychologische constructen: attitude ten aanzien van de statistiek (*Survey of Attitudes toward Statistics* (SATS), Schau, Stevens, Dauphinee en Del Vecchio, 1995), motivatie en leerstrategieën (*Motivated Strategies for Learning Questionnaire* (MSLQ), Pintrich, Smith, Garcia & McKeachie, 1991) en autonomie en afhankelijkheid (gebaseerd op de Five-Factor Personality Inventory, 'Big Five', Goldberg, 1990). De statistiekkennistest bestaat uit een aantal statistische opdrachten waarvan de antwoorden zijn geclassificeerd volgens de SOLO-taxonomie (*Structure of the Observed Learning Outcomes*) van Biggs en Collis (1982).

### *Procedure*

Voor het uitvoeren van Raschmodelanalyses is gebruik gemaakt van het softwareprogramma Winsteps (versie 5, Linacre, 2005). Dit programma voorziet in de volgende



criteria waarmee kan worden beoordeeld in hoeverre de Likertschaalitems respectievelijk de Likertschalen passen volgens het Raschmodel.

– *Fitmaten: Infit en Outfit*

Infit (information-weighted fit statistic) en Outfit (outlier-sensitive fit statistic) zijn maten die aangeven in hoeverre items geacht worden gezamenlijk een construct te meten. Er vindt een toetsing plaats van de assumptie dat de items in een hiërarchisch patroon passen waar items en respondenten op één enkel onderliggend continuüm – lopend van heel gemakkelijk tot heel moeilijk, of van geringe bekwaamheid tot grote bekwaamheid, of van geringe populariteit tot grote populariteit – geplaatst kunnen worden. Per item en per schaal voorziet Winsteps in chikwadratgroottheden die de mate van overeenstemming tussen de volgens het Raschmodel verwachte scores en de waargenomen scores weergeven. De Infit is gevoelig voor inconsistente of onverwachte antwoorden op items die in de nabijheid liggen van iemands positie op de schaal en is ongevoelig voor inconsistente antwoorden op items die ver van de bekwaamheden van een respondent staan. De Outfit daarentegen is gevoelig voor inconsistente antwoorden op items die ver van de bekwaamheden van een respondent staan en is ongevoelig voor inconsistente antwoorden op items die in de nabijheid liggen van iemands positie op de schaal liggen. Infit en Outfit maten worden gerapporteerd in diverse eenheden. In ongestandaardiseerde vorm worden ze gerapporteerd als Mean Squares in de vorm van chikwadratgroottheden gedeeld door hun vrijheidsgraden. De eenheden zijn van ratiomeetniveau met een verwachte waarde van +1 en een spreiding van 0 tot oneindig. Een Infit of Outfit Mean Square van  $1 + x$  geeft aan dat er  $100x\%$  meer variatie aanwezig is tussen de waargenomen en de door het model voorspelde antwoordpatronen. Daarnaast rapporteert Winsteps de Infit en Outfit ook in een bij benadering normaalverdeelde  $t$ -verdeling met verwachtingswaarde 0 en een standaardafwijking ongeveer gelijk aan 1 (Wilson-Hilferty transformatie, Bond en Fox, 2001, p. 177). Wright (in Linacre, 2005) wijst er overigens op dat de  $t$  “is only useful to salvage non-significant Mean Square  $> 1.5$ , when sample size is small or test length is short” (p. 198), waarmee hij wil aangeven dat deze fitmaat vooral nuttig is voor nader onderzoek van misfittende items.  $t$ 's  $< 0$  wijzen op gebrek aan waarschijnlijkheid in de data,  $t$ -waarden  $> 0$  op variabiliteit, wat nader onderzoek van mogelijke afwijkingen van unidimensionaliteit kan ondersteunen (Smith, 1994, in Linacre, 2005). Onderzoek heeft duidelijk gemaakt dat een type I fout van 5% wordt benaderd door een criterium van  $t \pm 2$  aan te houden. In dit proefschrift is het criterium gehanteerd dat elke gestandaardiseerde Infit of Outfit statistiek groter dan of gelijk moet zijn aan .5 en niet groter dan 1.5. Alle waarden buiten dit bereik worden beschouwd als niet passend in het Raschmodel (Curda, 1997; Linacre, 2005). In tabel 1 worden de criteria voor de Infit en Outfit genoemd die gebruikelijk zijn bij Raschmodelanalyses van oorspronkelijke Likertschalen (Bond & Fox, 2001).

Tabel 1 Raschmodel fitmaten voor Likertitems respectievelijk Likertschalen en hun algemene interpretatie (naar Bond & Fox, 2001)

Mean Squares	t	Antwoordpatroon	Variatie	Interpretatie	Misfit type
> 1.5	> 2.0	te toevallig	te veel	onvoorspelbaar, mogelijk afwijking van eendimensionaliteit, redundantie	underfit
< 0.5	< 2.0	te gedetermineerd	te weinig	Guttman	overfit

– *Antwoordpatronen: Average measures en Step calibration measures*

Winsteps voorziet in criteria waarmee kan worden beoordeeld in hoeverre de antwoordpatronen van de respondenten op een acceptabele, voorspelbare manier voldoen aan de door de hiërarchie van antwoorden verwachte antwoordpatronen. Er zijn allerlei redenen te bedenken waarom respondenten inconsequent lijken te reageren op items (raden, vermoeidheid, misinterpretatie van de richting, item meet nog iets anders dan betreffende variabele) waardoor een item niet of nog iets anders meet dan de variabele in kwestie. Bij inconsequenties kan de onderzoeker er niet meer van uitgaan dat de totale ruwe score voor alle personen op alle items de totale meting op de schaal weerspiegelt. Het Raschmodel voorziet in schattingen voor zowel elk item als voor elk individu op het continuüm van de onderliggende variabele. Winsteps vermeldt twee soorten criteria:

– *Average measures* houden rekening met de vorm van de verdeling en het aantal antwoorden per antwoordcategorie van de items. Het criterium voor goed passen is dat de antwoordcategorieën oplopen in (logit)waarde (monotoon stijgend) (Bond & Fox, 2001, p. 162).

– *Step calibration measures* geven aan in hoeverre de antwoordcategorieën van elkaar te onderscheiden zijn. Bond en Fox (2001, p. 163) geven als richtlijn dat de drempels (thresholds) 1) in waarde oplopen (monotoon stijgend) en 2) dat voor een voldoende onderscheid tussen de categorieën minstens 1.4 logits van elkaar moeten verschillen, maar niet meer dan 5 logits om grote gaten in de variabelenwaarden te vermijden (Bond & Fox, 2001, p. 163; Linacre, 2005).

– *Moelijkheidsgraad van de items op de schaal: Item Separation Index en Item Difficulty Scale*

Het Raschmodel transformeert de ruwe scores op de Likertitems via natuurlijke logaritmes van de odds kansen naar metrische logits. In de praktijk variëren de logits van ongeveer - 4 tot + 4, te interpreteren als van gemakkelijk tot moeilijk (respectievelijk van gematigdheid tot extreem, of van impopulair tot populair). De moeilijkheidsgraden (*measures*) geven aan welke positie een item inneemt in een schaal. Om na te gaan of de items voldoende verspreid over het onderliggend continuüm van de schaal liggen, bevelen Bond en Fox (2001) voor een minimale dekking minstens twee strata aan. Dit wordt aangegeven met de *Item Separation Index* (Bond & Fox, 2001, p. 207). In de praktijk hanteert men een iets minder stringent onderscheid, te weten: de Item Difficulty Scale (Kilgore, Fisher, Silverstein, Harley,

& Harvey, in L.K. Curda, 1997). Deze wordt verkregen door de Item Separation Index te transformeren volgens de formule  $[(4G+1)/3]$ , waarbij G staat voor de Item Separation Index. De Item Difficulty Scale moet minstens 2 bedragen om de centra van die strata meer dan toevallig van elkaar in moeilijkheidsgraad te kunnen onderscheiden.

- *Interitemcorrelaties: PTMEA (Point-Measure Correlation)*  
Winsteps genereert puntbiseriële correlatiecoëfficiënten tussen de antwoorden op een bepaald item met de totale score op de overige items van de betrokken personen c.q. de antwoorden van een persoon op de overige items (of van een persoon) met de antwoorden. De coëfficiënten moeten positief zijn.
- *Betrouwbaarheidsindices: Person Reliability Index, Item Reliability Index*  
In het Rasch-model wordt de betrouwbaarheid van de schaal geschat voor zowel personen als items:
  - De *Person Reliability Index* geeft aan hoe goed personen kunnen worden onderscheiden op de gemeten variabele. Met andere woorden, geschat wordt de herhaalbaarheid van iemands plaatsing over de andere items die hetzelfde psychologische construct meten. De berekeningswijze van de schatting is gebaseerd op het zelfde concept als Cronbach's alpha (het percentage waargenomen variantie in de antwoorden dat reproduceerbaar is). De waarden lopen van 0 tot 1 (Bond & Fox, 2001, p. 207). Voor Likertschalen is het gebruikelijk een minimale waarde van .70 te hanteren (Loewenthal, 2001, p. 12). De Person Reliability Index valt doorgaans wat lager uit dan Cronbach's alpha omdat ze de extremen niet meeneemt in haar berekeningen.
  - De *Item Reliability Index* geeft aan hoe goed items van elkaar onderscheiden kunnen worden op de gemeten variabele. Ook de waarden in deze schaal lopen van 0 tot 1. Voor de itembetrouwbaarheid wordt in dit proefschrift een waarde van minstens .95 aangehouden.
- *Vaardigheid van de personen: Person Separation Index en Person Ability Scale*  
De Person Separation Index geeft net als de Item Separation Index (zie hiervoor) het discriminerend vermogen van een schaal aan, nu niet op het terrein van de items in gemakkelijk en moeilijk, maar van personen met hoge en lage bekwaamheden. In dit onderzoek wordt echte niet de door Winsteps gegenereerde *Person Separation Index* gehanteerd, maar de *Person Ability Scale*. Deze index ontstaat na de transformatie met  $[(4G+1)/3]$  waarbij G staat voor de Separatie Index (zie ook hiervoor bij de Item Separation Index). De Person Ability Scale dient minstens 2 te bedragen.

Tabel 2 bevat een samenvatting van de te hanteren criteria voor de beoordeling van de fit van de items en de schalen. Deze criteria dienen simultaan te worden toegepast, waarbij voor de beoordeling van items en schalen in 'twijfelgevallen' of 'grensgebieden' de Infit- en Outfit-criteria het zwaarst wegen, gevolgd door de categorie-indices en de diverse betrouwbaarheden en als laatste de separatie-indices.

Tabel 2 Criteria voor de beoordeling van voor het Raschmodel acceptabele Likertitems en Likertschalen

<i>Criterium</i>	<i>Acceptabel indien</i>	<i>Prioriteit</i>
Infit Mean Square	tussen 0.5 - 1.5	1
Outfit mean Square	tussen 0.5 - 1.5	1
Average Measure	monotoon stijgend	2
Step Calibration Measure	monotoon stijgend én > 1.4 logits verschil tussen de drempels	2
Person Reliability Index	.60	2
Person Ability Scale	2	3
Item Reliability Index	.95	2
Item Difficulty Scale	2	3
Cronbach's alpha	.70	n.v.t.

### *Analyseschema*

Eerst worden alle schalen volgens de Likertschaalprocedure getoetst op interne consistentie en wordt met confirmatorische factoranalyse (CFA) gecontroleerd in hoeverre de items passen binnen een schaal. Vervolgens wordt er een Rasch-analyse uitgevoerd en wordt bekeken in hoeverre items en schaal voldoen aan de criteria van tabel 2. Bij niet-goedpassende items en/of schalen wordt voortdurend heen en weer geswitcht in bovengenoemde analyses om te achterhalen waarom items en/of schalen niet passen. Hierbij wordt ook Structural Equation Modeling (SEM) toegepast om te achterhalen in hoeverre het niet goed passen moet worden toegeschreven aan bijvoorbeeld cross-loadings.

### *Resultaten*

In deze sectie wordt per schaal eerst de bedoelde, oorspronkelijke schaal beoordeeld op de in tabel 2 vermelde criteria voor het Raschmodel. In het geval er niet voldaan is aan een of meer criteria wordt aangegeven welke stappen zijn ondernomen om de schaal toch passend te krijgen. Het resultaat van deze operatie wordt in een tweede tabel ontvouwd. Elke schaal eindigt met een figuur – de map van personen en items – waarin de respondenten en de schaalitems op een continuüm zijn gepositioneerd.

### *Attitude*

De *Survey of Attitudes toward Statistics* (SATS) van Schau, Stevens, Dauphinee en Del Vecchio (1995) bevat vier schalen: Affect, Waarde van statistiek, Cognitieve competentie en Moeilijkheid (zie tabel 5.3 in dit proefschrift voor nadere omschrijvingen).

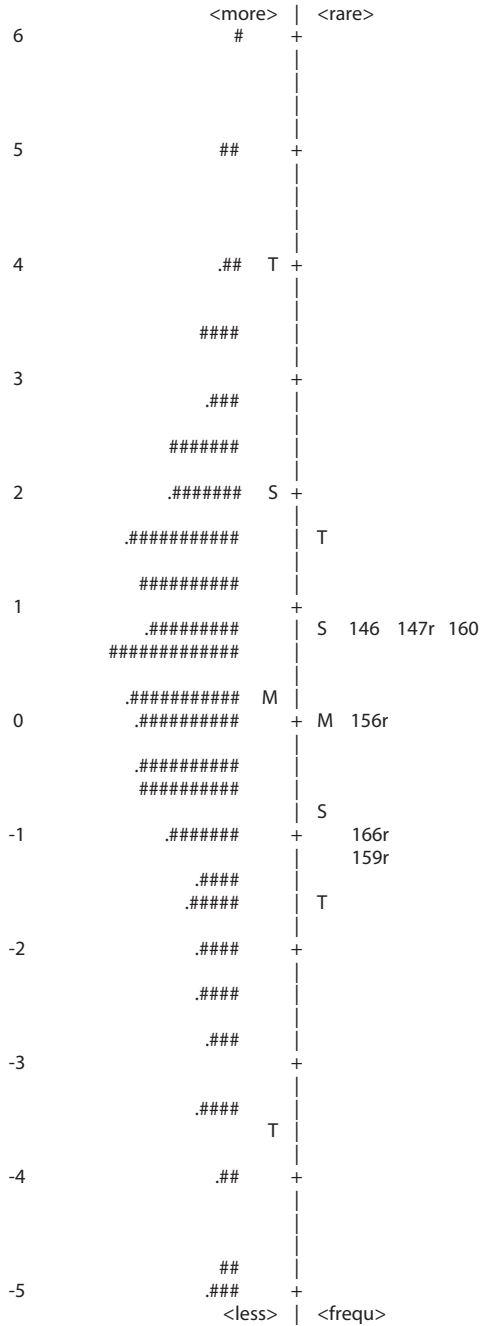
### *Affect*

In tabel 3 zijn de resultaten van de Raschanalyse vermeldt. De itemnummers verwijzen naar de vragenlijst in Bijlage II.2. Een r achter een itemnummer geeft aan dat de waarden van het oorspronkelijk negatief geformuleerde item zijn gehercodeerd ('reversed').

Tabel 3      *Affect*

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
147r	.96	1.04	.81	.07	.77		
146	1.05	1.02	.75	.07	.77		
160	1.05	1.12	.70	.07	.76		
156r	.74	.75	-.02	.07	.85		
166r	1.08	1.01	-.96	.07	.81		
159r	1.16	1.10	-1.29	.07	.77		
All items							
Mean	1.01	1.01	.00	.07		Person Reliability	.85
SD	.13	.12	.84	.00		Person Ability	3.53
All persons						Item Reliability	.99
Mean	.99	1.01	.19	.69		Item Difficulty	16.04
SD	.90	.94	1.86	.18		Cronbach's alpha	.89
Average measures	1	2	3	4	5		
	-3.04	-1.23	.05	1.36	2.98		
Step calibration measures		-2.98	-.72	.56	3.14		

De zes items van de schaal *Affect* voldoen op de Step calibration measures na alle aan de criteria van tabel 2. De afstand in logits van  $-.72$  naar  $.52$  in de drempels is kleiner dan 1.4. Gezien de prioriteitswaarde van deze maat zijn er geen redenen om deze schaal te reviseren. Figuur 2 geeft de door Winsteps gegenereerde Map van personen en items van deze schaal. Aan de linkerkant van de verticale lijn die het onderliggende continuüm van de latente trek representeert, zijn de respondenten geplaatst (elk #-teken staat in deze map voor 3 respondenten), aan de rechterkant de items. M staat voor respectievelijk Gemiddelde (bij de items staat deze op 0), S voor één standaardafwijking en T voor tweemaal de standaardafwijking in logits. De figuur toont dat zowel respondenten als items verspreid liggen rondom het gemiddelde. De scores van de respondenten lopen uiteen van  $-5$  tot  $+6$  logits. De items van ongeveer  $-1.30$  tot ongeveer  $+1$  (zie voor de precieze waarden de Measures in tabel 3).



EACH '#' IS 3.

Figuur 2 Affect: Map van personen items

*Waarde van statistiek*

Tabel 4a geeft de resultaten van de Raschanalyse van de negen oorspronkelijke items.

Tabel 4a Waarde van statistiek origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
158	.83	.85	1.34	.06	.72	*Item 161: In- and Outfit vallen buiten de range tot 1.5 er zijn inconsistenties in de categorieën: 3-2-1-4-5
153	1.06	1.14	1.25	.06	.59	
157r	.82	.83	.55	.06	.76	
152	1.28	1.27	.12	.06	.62	
164r	.74	.71	-.03	.07	.75	
170r	.62	.61	-.05	.07	.78	**Item 150: Toont te hoge Outfit, te wijten aan categorie 1 met een Outfit van 3.0
161r*	1.62	1.70	-.57	.07	.37	
155r	.86	.85	-1.08	.07	.70	
150r**	1.39	1.44	-1.52	.08	.59	
All items						
Mean	1.02	1.04	.00	.07		Person Reliability .81
SD	.32	.34	.91	.01		Person Ability 3.10
All persons						Item Reliability .99
Mean	1.03	1.04	.67	.55		Item Difficulty 17.10
SD	.84	.92	1.30	.15		Cronbach's alpha .84
Average measures	1	2	3	4	5	
	-1.76	-.75	.06	1.23	2.74	
Step calibration measures		-2.22	-1.14	.36	3.00	

Tabel 4a laat zien dat de oorspronkelijke schaal Waarde van statistiek niet geheel voldoet aan de Raschmodelcriteria. In het bijzonder is item 161 problematisch. Zowel de Infit- als de Outfitmaat zijn hoger dan de maximale waarde van 1.5 hetgeen wijst op te grote, op toeval beruste variatie van antwoorden. Reden om dit item te verwijderen uit de subschaal. Ook item 150, dat met zijn Infit- en Outfitmaten zelf nog valt binnen de toelaatbare grenzen, is problematisch gezien een Outfit van 3.0 voor antwoordcategorie 1. In combinatie met de bevinding met de ruwe Likertschaalscores in CFA en in SEM dat dit item blijkt te interfereren (cross-loadings) met andere items, is besloten dit item ook uit de subschaal te verwijderen. Tabel 4b toont de gereviseerde subschaal van Waarde van statistiek.

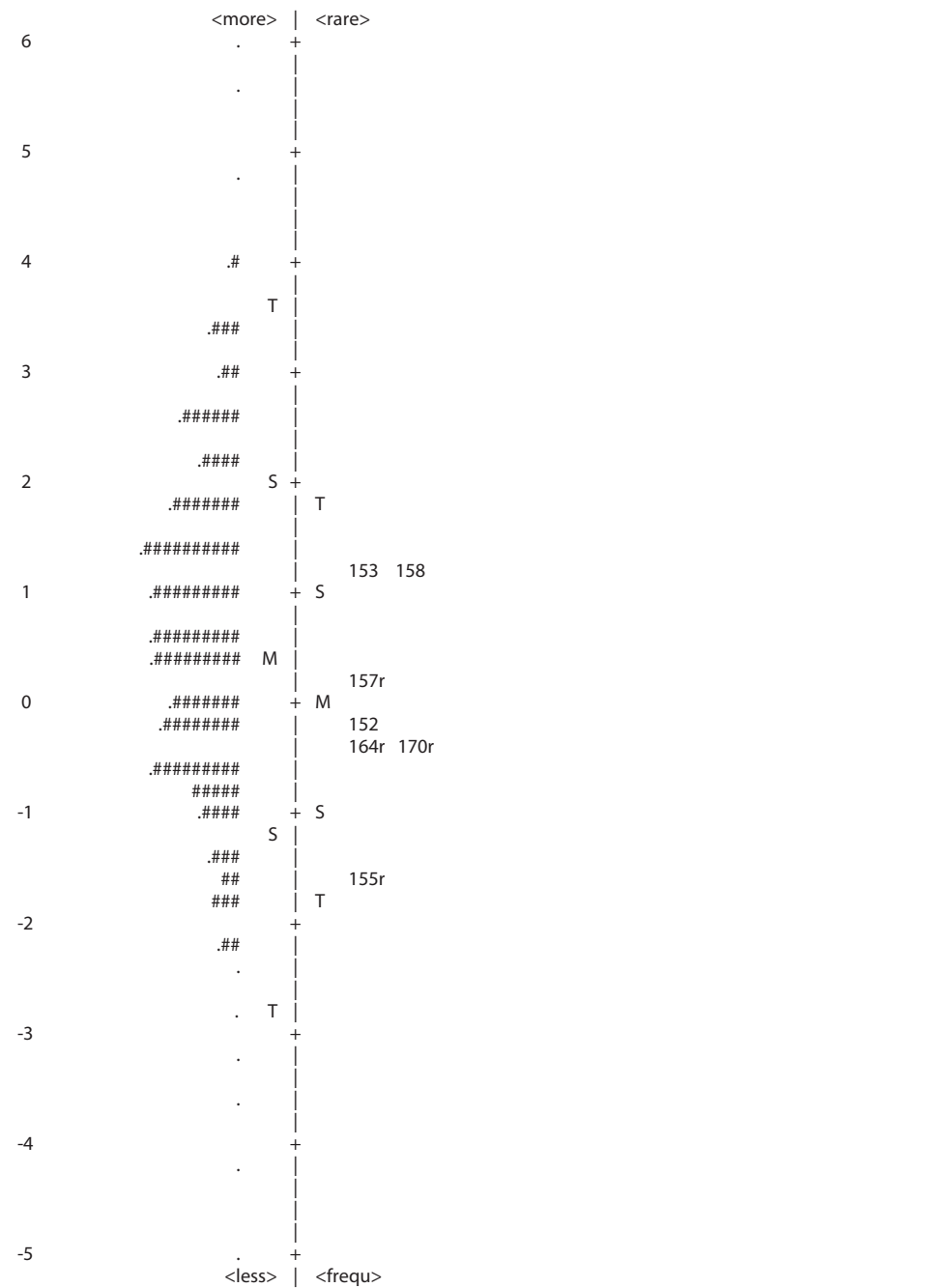
Tabel 4b Waarde van statistiek gereviseerd (SATS, Schau et al., 1995)

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
158	.96	.98	1.24	.07	.73		
153	1.21	1.28	1.14	.07	.62		
157r	.94	.94	.30	.07	.77		
152	1.44	1.44	-.22	.07	.65		
164r	.82	.77	-.40	.07	.77		
170r	.72	.69	-.42	.07	.79		
155r	.95	.92	-1.64	.08	.72		
All items							
Mean	1.01	1.00	.00	.07		Person Reliability	.82
SD	.23	.25	.93	.01		Person Ability	3.20
All persons						Item Reliability	.99
Mean	1.01	1.00	.47	.66		Item Difficulty	17.10
SD	.87	.88	1.61	.17		Cronbach's alpha	.85
Average measures	1	2	3	4	5		
	-2.56	-1.12	-.06	1.37	3.19		
Step calibration measures		-2.77	-1.14	.56	3.54		

Na hercalibratie van de data voldoet de gereviseerde subschaal aan de criteria van het Raschmodel.

Figuur 3 toont de Map van personen en items. Gezien de positionering van personen en items lijken de items enigszins aan de 'gemakkelijke' kant te liggen.





EACH '#' IS 4.

Figuur 3 Waarde van statistiek gereviseerd: Map van personen en items

## Cognitieve competentie

Tabel 5a geeft de resultaten van de Raschanalyse van de zes oorspronkelijke items.

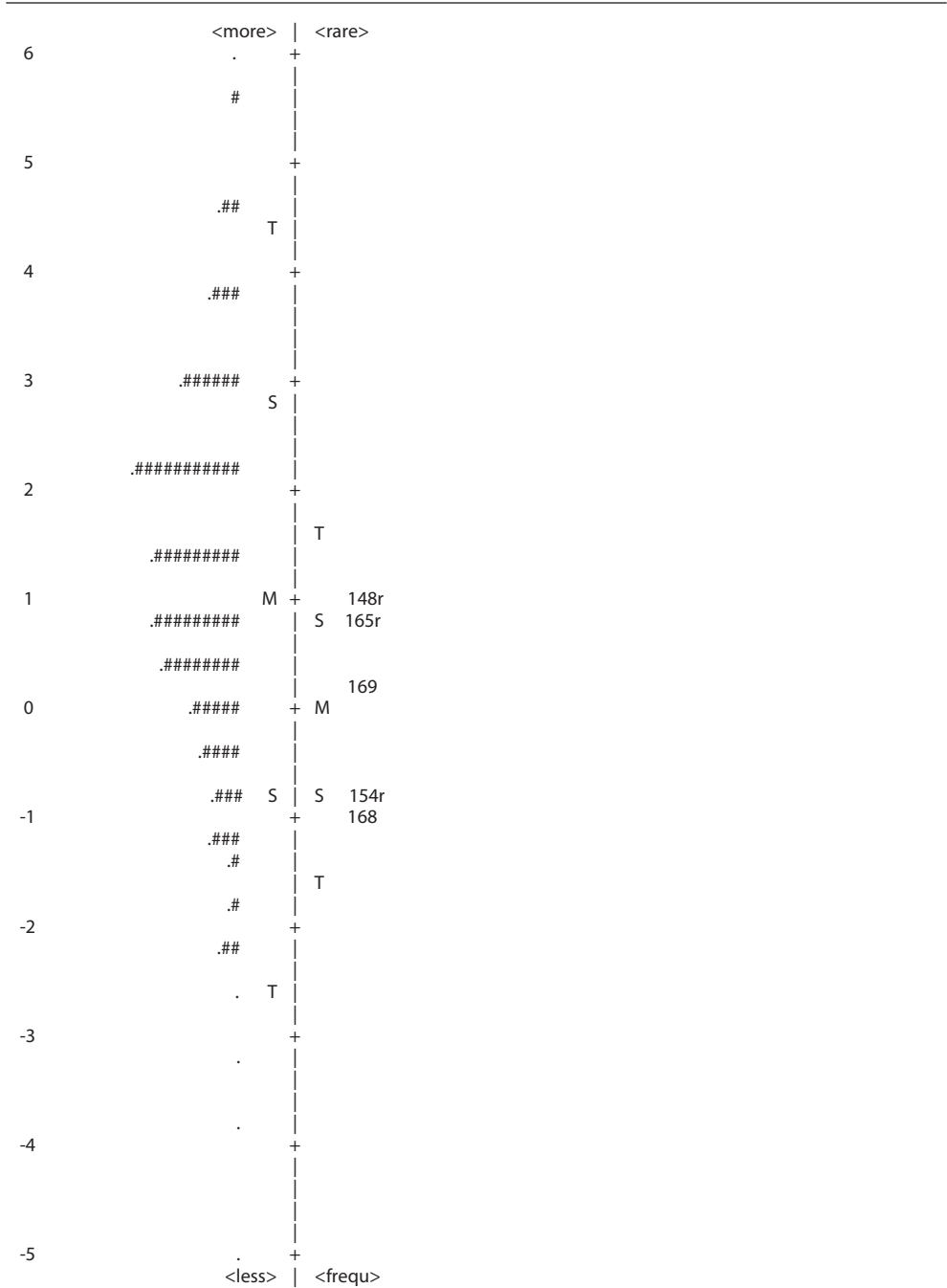
Tabel 5a Cognitieve competentie origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
172r	.90	.92	1.16	.07	.76	
148r	.83	.84	.67	.07	.81	
165r	1.29	1.37	.55	.07	.68	
169	.68	.68	.03	.07	.78	
154r	1.47	1.39	-1.10	.08	.69	
168	.87	.82	-1.30	.08	.72	
All items						
Mean	1.01	1.00	.00	.08		Person Reliability .80
SD	.28	.27	.91	.01		Person Ability 2.96
All persons						Item Reliability .99
Mean	1.00	1.00	.73	.74		Item Difficulty 15.72
SD	.91	.94	1.69	.20		Cronbach's alpha .85
Average measures	1	2	3	4	5	
	-2.58	-1.12	-.04	1.54	3.32	
Step calibration measures		-3.01	-1.04	.11	3.94	

Hoewel item 172 binnen de toelaatbare grenzen van Infit en Outfit scoort ( $> 0.5$ ), blijkt uit SEM-analyses met de SATS dat dit item interfereert met andere factoren. Op de factor Cognitieve competentie laadt dit item onvoldoende ( $\beta = .16$ ,  $p = .03$ ) en hoog op andere factoren zoals Moeilijkheid ( $\beta = -.72$ ,  $p = .00$ ). Een Raschanalyse met dit item in de schaal Moeilijkheid laat zien dat dit item ook daar binnen de Raschmodelcriteria zou passen (Infit en Outfit: .76 and .77; Measure: .84; SE: .06; PTMEA: .77). Gezien deze dubbelzinnigheid is besloten dit item te verwijderen. Tabel 5b toont de gereviseerde subschaal. Figuur 4 toont de Map van personen en items. In deze figuur blijkt dat het 'moeilijkste' item (item 148r) op hetzelfde gemiddelde ligt als de gemiddelde score van de respondenten (0.93 logits). Dit wijst erop dat deze schaal niet het 'moeilijkere' deel van het continuüm van de Cognitieve competentie meet.

Tabel 5b      Cognitieve competentie gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
148r	.82	.81	.93	.07	.82		
165r	1.36	1.45	.80	.07	.68		
169	.66	.68	.27	.07	.79		
154r	1.38	1.32	-.90	.08	.72		
168	.82	.75	-1.09	.08	.74		
All items							
Mean	1.01	1.00	.00	.08		Person Reliability	.76
SD	.30	.32	.84	.01		Person Ability	2.69
All persons						Item Reliability	.99
Mean	.99	1.00	.93	.84		Item Difficulty	14.09
SD	.95	.99	1.78	.24		Cronbach's alpha	.82
Average measures	1	2	3	4	5		
	-2.52	-1.11	-.05	1.64	3.54		
Step calibration measures		-2.87	-1.21	.03	4.05		



EACH '#' IS 6.

Figuur 4      Cognitieve competentie gereviseerd: Map van personen en items

**Moeilijkheid**

Tabel 6a geeft de resultaten van de Raschanalyse van de zeven oorspronkelijke items weer.

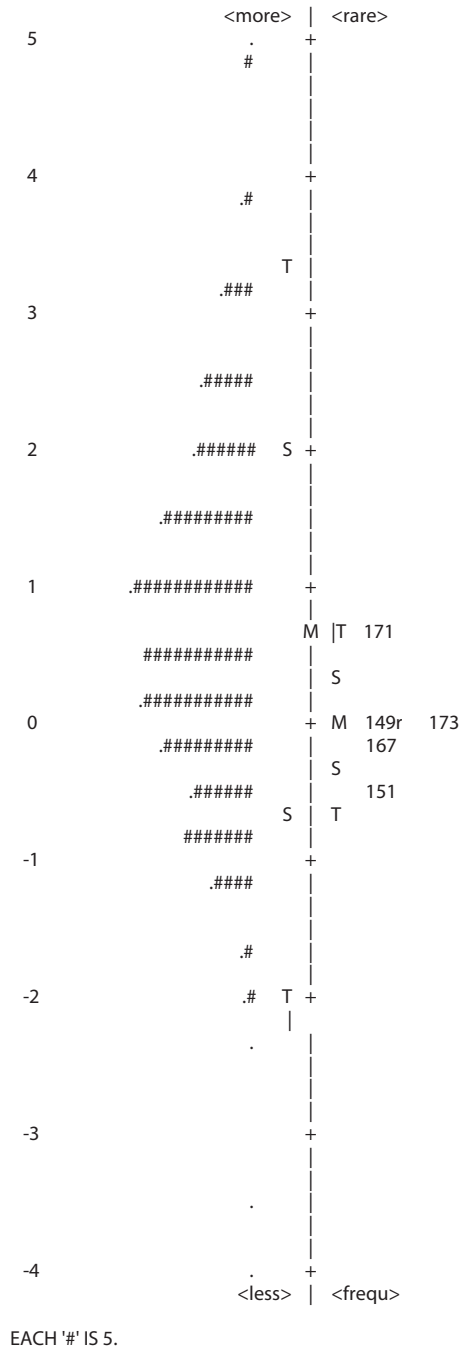
**Tabel 6a** Moeilijkheid origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
171	.94	.98	.96	.06	.64	*Item 163: Toont inconsistenties in de categorieën 2-1-3-4-5
173	.86	.89	.35	.07	.63	
149r	1.10	1.14	.34	.07	.64	
167	1.20	1.22	.27	.07	.59	
151	.99	1.00	-.07	.07	.71	** Item 162: Categorie 2 heeft te lage Outfit
163*	.91	.91	-.58	.07	.67	
162r**	.95	1.02	-1.27	.08	.51	
All items						
Mean	.99	1.02	.00	.07		Person Reliability .73
SD	.11	.11	.68	.00		Person Ability 2.52
All persons						Item Reliability .99
Mean	1.03	1.02	.93	.64		Item Difficulty 13.13
SD	.88	.88	1.28	.18		Cronbach's alpha .76
Average measures	1	2	3	4	5	
	-.93	-.47	.24	1.32	2.85	
Step calibration measures		-.265	-.97	.39	3.22	

Op het eerste gezicht lijkt de schaal Moeilijkheid grotendeels te voldoen aan de Raschmodelcriteria. Inspectie van de Infit en de Outfit van de afzonderlijke items laten zien dat de antwoordcategorieën van item 163 misfits in de ordening vertonen en dat antwoordcategorie 2 van item 162 een te lage Outfit kent (0.1). Beide items worden om die reden uit de schaal verwijderd. Tabel 6b toont de Raschmodeluitkomsten van de gereviseerde schaal en figuur 5 de Map van personen en items. De moeilijkheidschaal blijft echter een zwakke schaal dat niet het moeilijkste deel van het continuüm beslaat. Het moeilijkste item (171) ligt op dezelfde positie (.61 logits) van het continuüm als de gemiddelde persoonsmaat.

**Tabel 6b** Moeilijkheid gerevisiseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
171	.94	.96	.61	.07	.67	Person Reliability .66 Person Ability 2.20 Item Reliability .96 Item Difficulty 6.88 Cronbach's alpha .71
173	.87	.90	-.02	.07	.65	
149r	1.07	1.11	-.03	.07	.67	
167	1.16	1.17	-.10	.07	.62	
151	.98	.96	-.46	.07	.73	
All items						
Mean	1.00	1.02	.00	.07		
SD	.10	.10	.35	.00		
All persons						
Mean	1.02	1.02	.61	.75		
SD	.95	.96	1.33	.20		
Average measures	1	2	3	4	5	
	-1.00	-.62	.11	1.15	2.73	
Step calibration measures		-2.78	-1.00	.48	3.31	



Figuur 5 Moeilijkheid gereviseerd: Map van personen en items

*Motivatie*

De *Motivated Strategies for Learning Questionnaire* (MSLQ, Pintrich, Smith, Garcia & McKeachie, 1991) bevat in het motivatiedeel vier schalen: Intrinsieke waarde, Taakwaardering, Zelfeffectiviteit en Testangst (zie tabel 5.4 voor nadere omschrijvingen).

*Intrinsieke waarde*

Tabel 7a geeft de resultaten van de Raschanalyse van de vijf oorspronkelijke items weer.

Tabel 7a Intrinsieke waarde origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
5	1.07	1.12	.39	.07	.60	
6	1.11	1.13	.39	.07	.60	
4	1.02	.98	-.21	.08	.65	
8	.79	.80	-.28	.08	.64	
76	.91	.93	-.30	.08	.49	
All items						
Mean	.98	.99	.00	.08		Person Reliability .48
SD	.12	.12	.32	.00		Person Ability 2.48
All persons						Item Reliability .94
Mean	.99	.99	1.69	.88		Item Difficulty 7.69
SD	.12	.12	1.26	.25		Cronbach's alpha .59
Average measures	1	2	3	4	5	
	-.56	.05	.78	1.66	3.02	
Step calibration measures		-2.54	-.77	-.18	3.50	

Hoewel de schaal Intrinsieke waarde voldoet aan de Infit en de Outfit voor zowel items als personen, blijken de Person Reliability Index (<.60), de Item Reliability Index (<.95) en de drempels in de Step calibration measures (<1.4) onvoldoende te zijn. De range van de items op het continuüm is zeer beperkt (lopend van -.30 tot +.39 logits) en het gemiddelde van de items ligt ver onder het gemiddelde van de personen. Dat laatste houdt in dat de items veel te 'gemakkelijk' zijn om te beantwoorden. De gemiddelde score van de personen op deze schaal (1.69 logits) is ruim viermaal hoger dan de score van het 'moeilijkste' item (item 5: .39 logits).

Het discriminerend vermogen van de subschaal is laag. Twee paren van de vijf items hebben ongeveer dezelfde moeilijkheidswaarden op het continuüm (item 5 en 6 met een waarde van .39 en de items 8 en 76 met waarden -.28 en -.30). Factoranalyse van de residuen genereert een tweetal onderliggende factoren voor de items 4 en 6 en de items 5, 76 en 8. Vanuit theoretisch oogpunt zou dit onderscheid geïnterpreteerd kunnen worden als een verschil in cognitieve intrinsieke waarde en extrinsieke cognitieve taakwaarde. Binnen de factor zelf zijn affectieve aspecten ondervertegenwoordigd. Aangezien bij de vastlegging van de vragenlijst dit probleem was voorzien, zijn er aanvullende items geformuleerd op basis van Flow Theory (Czikszentmihalyi, 1990)

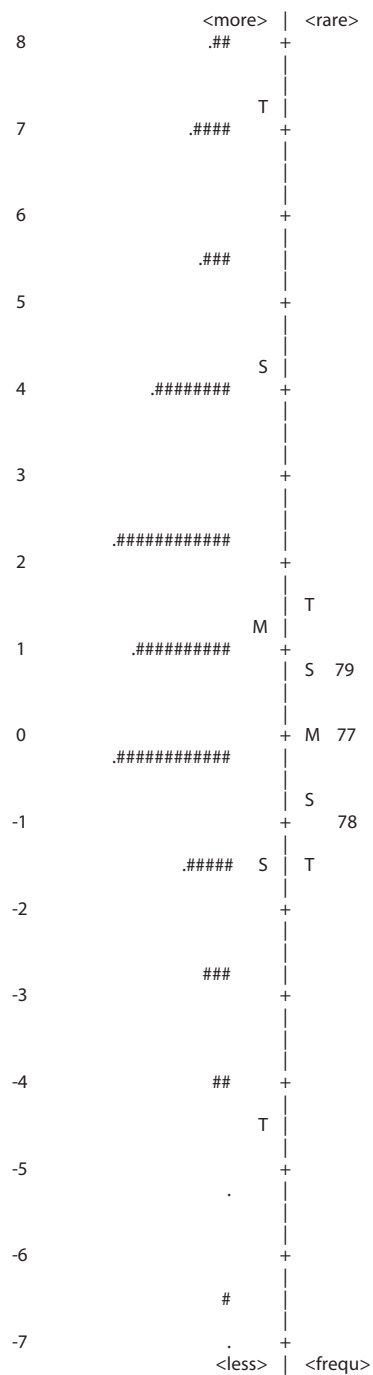
die als de items 77, 78 en 79 in de vragenlijst zijn opgenomen. Deze drie items leveren de volgende Raschmodeluitkomsten op (tabel 7b).

Tabel 7b Intrinsieke waarde gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
79	.80	.78	.80	.09	.89		
77	1.34	1.31	.09	.10	.77		
78	.83	.78	-.89	.10	.89		
All items							
Mean	.99	.96	.00	.10		Person Reliability	.78
SD	.25	.25	.69	.01		Person Ability	2.84
All persons						Item Reliability	.98
Mean	.97	.96	1.36	1.34		Item Difficulty	9.33
SD	1.13	1.16	2.96	.36		Cronbach's alpha	.82
Average measures	1	2	3	4	5		
	-5.41	-2.33	.17	3.02	6.34		
Step calibration measures		-5.65	-2.16	1.51	6.30		

De drie vervangende items blijken beter aan de Raschmodelvereisten te voldoen dan de oorspronkelijke vijf items uit de MSLQ. De variatiebreedte van waarden van de items is iets meer uitgebreid (van -.30 naar -.89 logits en van .39 naar .80 logits) en de gemiddelde persoonsmaat scoort nog maar 1.7 maal hoger dan het moeilijkste item (item 79). Dit laatste houdt in dat in vergelijking met de vorige subschaal de items een stuk naar het midden van het continuüm zijn opgeschoven. De items blijven nog wel aan de 'gemakkelijke' kant. Figuur 6 toont de verdeling van persoonsscores en itemposities op het onderliggend continuüm van de subschaal intrinsieke waarde. De drie verwijderde items uit de originele subschaal die qua inhoud beter pasten bij Taakwaarde, blijken bij nadere analyse inderdaad goed te passen bij deze. Deze items (5, 8 en 76) zijn in die schaal opgenomen (zie hierna).





EACH '#' IS 7.

Figuur 6 Intrinsieke waarde gereviseerd: Map van personen en items

## Taakwaardering

Tabel 8a geeft de resultaten van de Raschanalyse van de zes oorspronkelijke items

Tabel 8a Taakwaardering origineel

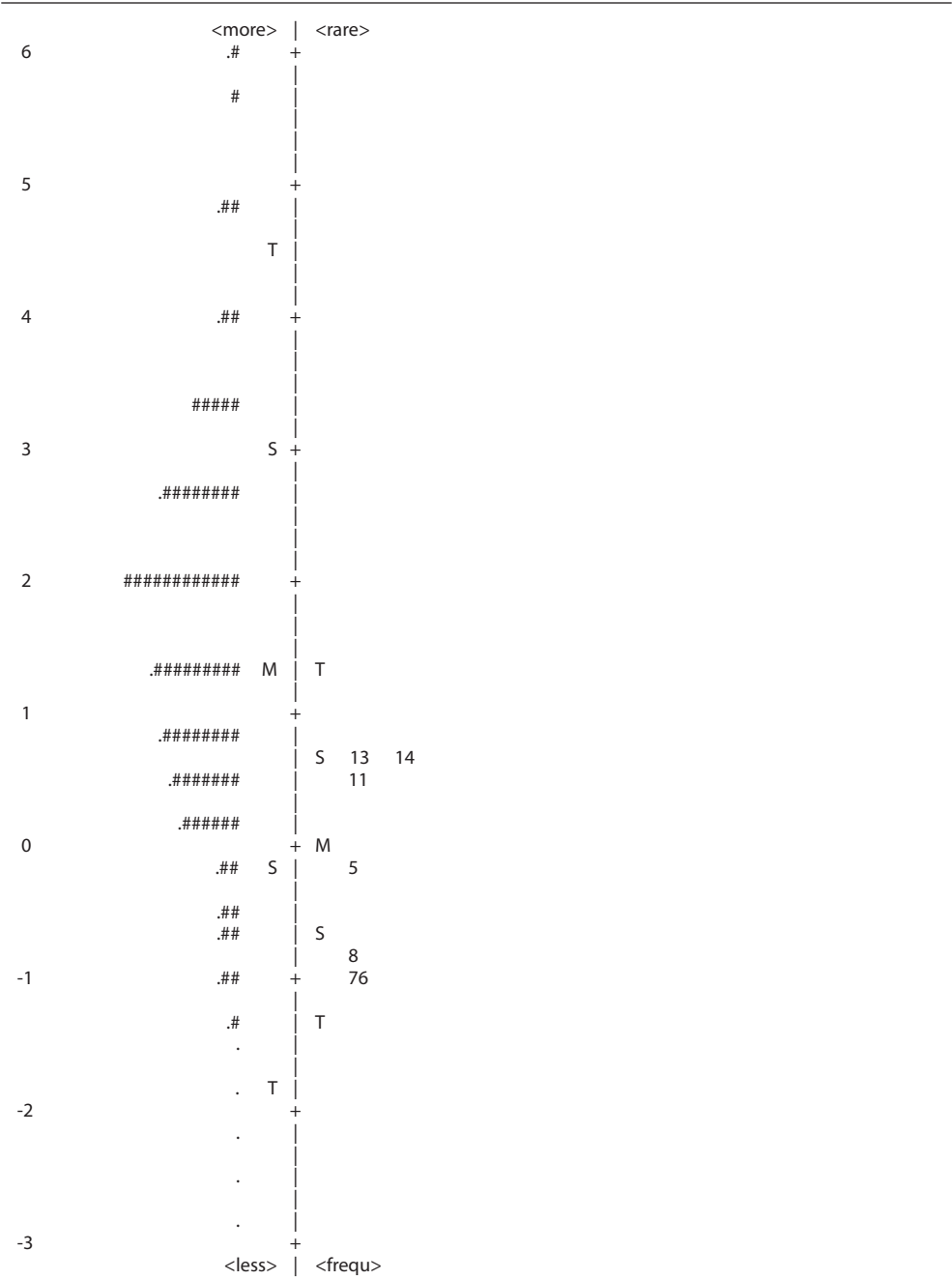
Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous
14	.77	.95	.77	.08	.74	
13	.74	.92	.74	.08	.78	
11	1.11	1.09	.55	.08	.73	
12	1.06	1.02	-.63	.09	.70	
10	.82	.75	-.69	.09	.78	
9	1.15	1.09	-.74	.09	.67	
All items						
Mean	1.00	.97	.00	.08		Person Reliability .80
SD	.12	.12	.69	.01		Person Ability 3.00
All persons						Item Reliability .98
Mean	1.51	.83	1.51	.83		Item Difficulty 11.13
SD	1.91	.22	1.91	.22		Cronbach's alpha .85
Average measures	1	2	3	4	5	
	-2.58	1.00	.20	1.94	4.15	
Step calibration measures		-3.33	-1.15	.02	4.46	

Tabel 8a laat zien dat de schaal Taakwaardering grotendeels aan de criteria van een Raschmodel voldoen, op de Persons Infit ( $< 1.5$ ) en de twee middelste drempels in de Step calibration measures ( $< 1.4$ ) na. Ook deze schaal blijkt relatief 'gemakkelijk' om te beantwoorden: het gemiddelde van alle personen (1.51) ligt bijna tweemaal boven het moeilijkste item (item 14: .77). Nadere analyses met CFA en SEM laten zien dat de items in deze subschaal interfereren met andere items en schalen. In combinatie met de oorspronkelijke Intrinsieke waardeschaal blijkt redundantie op te treden met enkele items (de items 5, 8 en 76). Nadat de oorspronkelijke Intrinsieke waardeschaal is vervangen door de gereviseerde Intrinsieke waardeschaal (Flow), blijken de cross-loadings te zijn verdwenen.

De items 5, 8 en 76 blijken goed te passen binnen de Taakwaardeschaal. Item 76 krijgt zelfs nog een functie als equivalentieitem ('anchoring') in het longitudinale onderzoek (zie hoofdstuk 6 van dit proefschrift). De items 9, 10 en 12 blijken met moeilijkheidswaarden die dicht bij elkaar liggen redundantie te vertonen. Om efficiëntieredenen is besloten deze items uit de schaal voor Taakwaardering te verwijderen. Het resultaat van deze toevoegingen en verwijderingen wordt getoond in tabel 8b. Vergeleken met de oorspronkelijke subschaal voor Taakwaardering passen in het gereviseerde model de Raschmodelcriteria nu beter, hoewel nog steeds niet overtuigend. Het blijft een zwakke schaal. Op face validity niveau voldoen de geselecteerde items beter aan de inhoudelijke betekenis van het beoogde construct dan de oorspronkelijke items. Figuur 7 toont de map van personen en items.

Tabel 8b      Taakwaardering gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
14	.97	.97	.73	.07	.69		
13	.73	.76	.70	.07	.79		
11	.89	.90	.54	.07	.74		
5	1.19	1.27	-.14	.08	.60		
8	.87	.85	-.90	.09	.65		
76	1.14	1.19	-.92	.09	.47		
All items							
Mean	.97	.99	.00	.08		Person Reliability	.72
SD	.16	.18	.70	.01		Person Ability	2.48
All persons						Item Reliability	.99
Mean	.97	.99	1.36	.79		Item Difficulty	11.89
SD	.90	.95	1.57	.24		Cronbach's alpha	.78
Average measures	1	2	3	4	5		
	-1.91	-.76	.26	1.68	3.39		
Step calibration measures		-2.69	-1.13	-.14	3.96		



EACH '#' IS 6.

Figuur 7 Taakwaardering gereviseerd: map van personen en items

**Zelfeffectiviteit**

Tabel 9a geeft de resultaten van de Raschanalyse van de twaalf oorspronkelijke items.

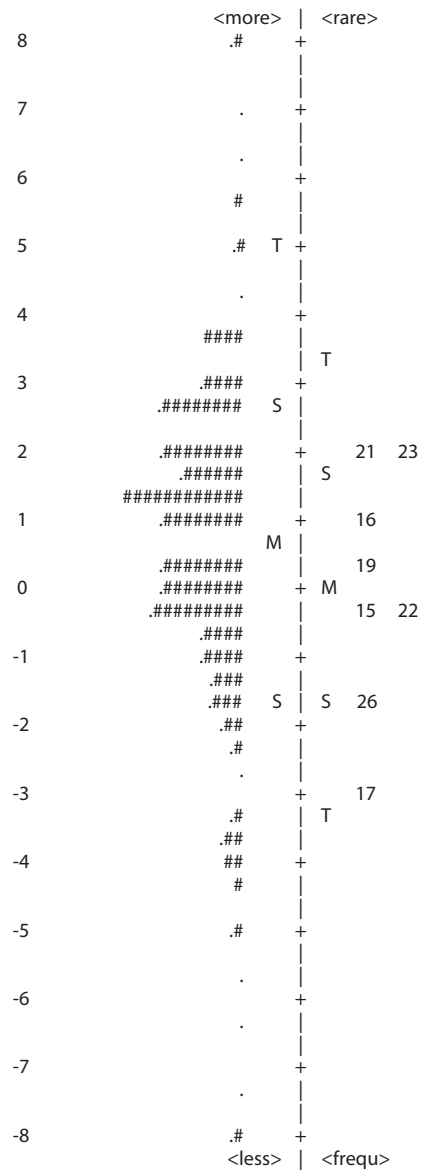
Tabel 9a      Zelfeffectiviteit origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
25	1.19	1.26	1.63	.08	.67	
24	.97	1.49	1.46	.08	.73	
21	.75	.74	1.40	.08	.81	
23	.87	.90	1.23	.08	.77	
18	1.11	1.09	.52	.08	.79	
16	1.20	1.21	.49	.08	.78	
20	1.02	1.04	.06	.08	.77	
19	.78	.77	-.15	.08	.82	
15	1.33	1.33	-.71	.08	.75	
22	.80	.78	-.75	.08	.82	
26	.84	.79	-2.03	.09	.76	
17	.99	.97	-3.13	.09	.73	
All items						
Mean	.99	1.03	.00	.08		Person Reliability .92
SD	.18	.24	1.41	.01		Person Ability 4.92
All persons						Item Reliability 1.00
Mean	1.01	1.01	.16	.56		Item Difficulty 22.83
SD	.80	.86	2.09	.15		Cronbach's alpha .94
Average measures	1	2	3	4	5	
	-4.14	-1.87	-.21	1.93	4.50	
Step calibration measures		-4.11	-1.55	.95	4.71	

De effectiviteitsschaal met twaalf items voldoet aan alle Raschmodelcriteria. Vanuit spaarzaamheidsoverwegingen zijn items die een ongeveer gelijke positie op het onderliggend continuüm bezetten, zoals dat in de clusters 19-20 en 21- 23- 24 het geval is, uit de schaal verwijderd. Nadere inspectie op face validity, Raschmodelcriteria voor de items en het gedrag in SEM hebben geleid tot verwijdering van de items 25, 24, 18, 20, 24 en 25. Tabel 9b toont de resterende schaal. Figuur 8 toont de map met personen en items.

Tabel 9b Zelfeffectiviteit gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
21	.83	.85	2.03	.08	.81	
23	1.01	1.11	1.85	.08	.77	
16	1.21	1.23	1.05	.08	.79	
19	.81	.82	.36	.08	.83	
15	1.34	1.40	-.26	.09	.77	
22	.83	.82	-.30	.09	.83	
26	.86	.80	-1.76	.09	.78	
17	1.00	.96	-2.98	.10	.75	
All items						
Mean	.99	1.00	.00	.09		Person Reliability .89
SD	.18	.21	1.62	.01		Person Ability 4.07
All persons						Item Reliability 1.00
Mean	1.00	1.00	.57	.56		Item Difficulty 24.56
SD	.81	.85	2.22	.15		Cronbach's alpha .92
Average measures	1	2	3	4	5	
	-4.48	-2.09	-.16	2.32	4.85	
Step calibration measures		-4.47	-1.68	.87	5.28	



EACH '#' IS 4.

Figuur 8 Zelfeffectiviteit gereviseerd: Map van personen en items

## Testangst

Tabel 10a geeft de resultaten van de Raschanalyse van de vier oorspronkelijke items.

Tabel 10a Testangst origineel

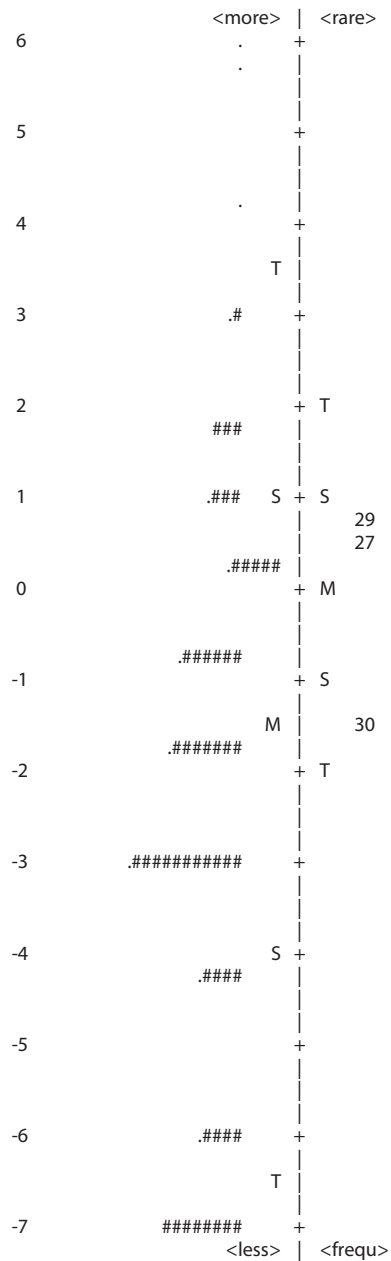
<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
29	.96	.98	.80	.09	.85		
27	.96	1.00	.54	.09	.85		
28	.75	.74	.12	.09	.90		
30	1.23	1.32	-1.47	.09	.86		
All items							
Mean	.98	1.01	.00	.09		Person Reliability	.80
SD	.17	.21	.88	.00		Person Ability	3.01
All persons						Item Reliability	.99
Mean	1.00	1.01	-1.59	1.08		Item Difficulty	12.76
SD	1.22	1.26	2.55	.35		Cronbach's alpha	.90
Average measures	1	2	3	4	5		
	-5.11	-2.51	-.30	1.63	4.23		
Step calibration measures		-4.96	-.64	1.02	4.59		

De schaal Testangst met vier items voldoet aan de Raschmodelcriteria. Vanuit spaarzaamheidsoverwegingen zou het voor Infit en Outfit iets meer gevoelige item 28 uit de subschaal verwijderd kunnen worden. Nadere inspectie van de items in deze subschaal met de andere items brengt aan het licht dat item 28 cross-loadings vertoont met andere items en factoren. Besloten is dit item uit de schaal te verwijderen. Tabel 10b toont de resulterende schaal. Figuur 9 geeft de map met personen en items weer. De schaal blijkt, net als de vier-item schaal, het moeilijke deel van het continuüm beter te meten dan het gemakkelijke deel. Het moeilijkste item (item 29, .85 logit) overtreft de gemiddelde persoonsscore op de testangst (-1.54 logit) 2.8 maal. Het gemakkelijkste item (item 30, -1.44 logit) overtreft eveneens de gemiddelde persoonsscore.

Tabel 10b Testangst gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
29	.89	.91	.85	.09	.86		
27	.92	.97	.59	.09	.86		
30	1.10	1.08	-1.44	.09	.88		
All items							
Mean	.97	.99	.00	.09		Person Reliability	.71
SD	.09	.07	1.02	.00		Person Ability	2.43
All persons						Item Reliability	.99
Mean	.99	.99	-1.54	1.26		Item Difficulty	14.87
SD	1.30	1.28	2.56	.41		Cronbach's alpha	.84
Average measures	1	2	3	4	5		
	-4.90	-2.58	-.35	1.73	4.24		
Step calibration measures		-5.04	-.67	.98	4.73		





EACH '#' IS 4.

Figuur 9 Testangst gereviseerd: Map van personen en items

*Leerstrategieën*

De *Motivated Strategies for Learning Questionnaire* (MSLQ, Pintrich et al., 1991) bevat in het leerstrategiegedeelte vijf schalen: Herhalen, Elaboreren, Organiseren, Kritisch denken en Metacognitie (zie tabel 5.5 voor nadere omschrijvingen).

## Herhalen

Tabel 11a geeft de resultaten van de Raschanalyse van de drie oorspronkelijke items weer.

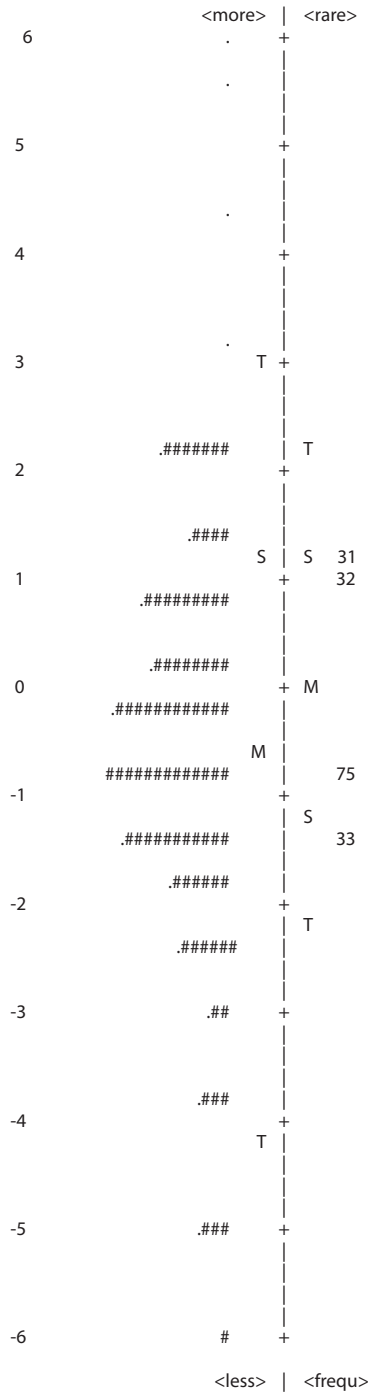
Tabel 11a Herhalen origineel (MSLQ, Pintrich et al, 1991)

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
31	.91	.92	1.03	.08	.80		
32	.67	.65	.92	.08	.84		
33	1.29	1.26	-1.94	.08	.74		
All items							
Mean	.96	.94	.00	.08		Person Reliability	.66
SD	.25	.25	1.38	.01		Person Ability	2.19
All persons						Item Reliability	1.00
Mean	.91	.94	-.78	1.10		Item Difficulty	22.57
SD	1.13	1.20	1.96	.30		Cronbach's alpha	.72
Average measures	1	2	3	4	5		
	-4.11	-2.10	-.53	1.58	2.24		
Step calibration measures		-4.05	-.76	.03	4.77		

De schaal Herhalen voldoet grotendeels aan de Raschmodelvereisten, in ieder geval wat betreft Infit en Outfit en de betrouwbaarheden voor items en personen en de monotoon stijgende waarden bij Average measures en Step calibration measures. Een punt van discussie vormt het verschil tussen de twee middelste drempels die lager is dan 1.4. Twee items hebben nagenoeg dezelfde positie op het onderliggende continuüm. Om het discriminerend vermogen van de schaal te verhogen, is een item (item 75) toegevoegd. Deze toevoeging leidt tot een schaal waar de vier items evenwichtiger over het gemakkelijke en moeilijke deel van het continuüm zijn verspreid. Zie figuur 10 voor een verdeling van de personen en items over het onderliggend continuüm van het construct Herhalen. De persoonsbetrouwbaarheid is eveneens gestegen (van .66 naar .73).

Tabel 11b Herhalen gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
31	.95	.94	1.12	.07	.77		
32	.70	.67	1.03	.07	.82		
75	1.01	1.06	-.73	.07	.73		
33	1.24	1.20	-1.42	.07	.70		
All items						Person Reliability	.73
Mean	.98	.97	.00	.07		Person Ability	2.53
SD	.19	.19	1.10	.01		Item Reliability	1.00
All persons						Item Difficulty	19.99
Mean	.97	.97	-.60	0.89		Cronbach's alpha	.76
SD	1.02	1.05	1.78	.25			
Average measures	1	2	3	4	5		
	-3.62	-1.72	-.43	1.22	2.16		
Step calibration measures		-3.53	-.69	-.04	4.26		



EACH '#' IS 5.

Figuur 10 Herhalen gereviseerd: map van personen en items

**Elaboreren**

Tabel 12a geeft de resultaten van de Raschanalyse van de vijf oorspronkelijke items weer.

Tabel 12a    Elaboreren origineel

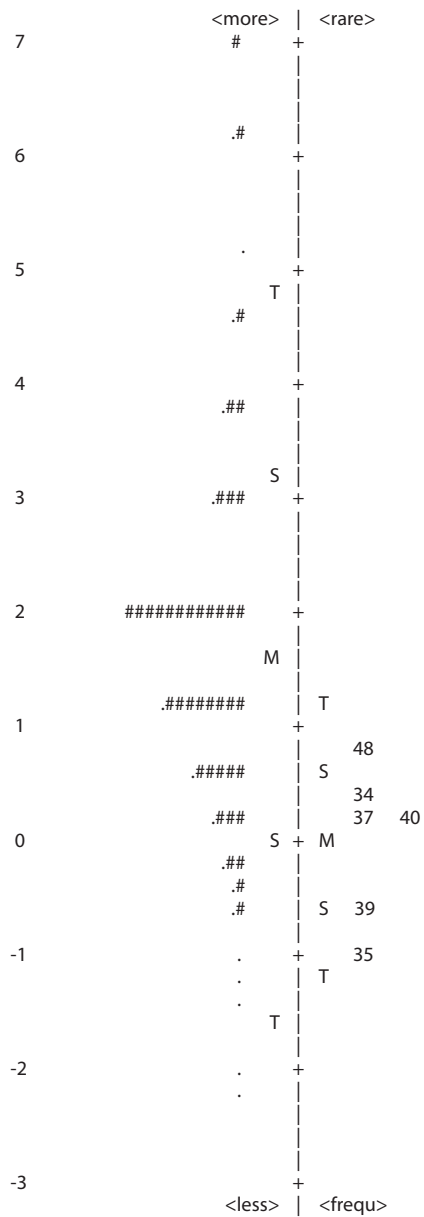
<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
34	.89	.84	.82	.09	.72	* Item 36 toont inconsistenties in de categorieën 2-3-1-4-5
37	.65	.55	.63	.10	.76	
36*	1.43	1.42	.30	.10	.65	
38	1.08	1.09	-.72	.12	.62	
35	.74	.59	-1.02	.12	.73	
All items						
Mean	.96	.90	.00	.11		Person Reliability .55
SD	.27	.32	.74	.01		Person Ability 1.81
All persons						Item Reliability .98
Mean	.85	.87	2.29	1.23		Item Difficulty 9.19
SD	1.30	1.37	1.99	.49		Cronbach's alpha .77
Average measures	1	2	3	4	5	
	-1.16	-1.15	.26	2.38	5.41	
Step calibration measures		-3.71	-1.48	-.54	5.73	

Tabel 12a toont dat de originele Elaboratieschaal niet voldoet aan de Raschmodelvereisten. Zowel Person Reliability Index ( $< .60$ ), Person Ability Scale ( $< 2$ ) en de afstand tussen de twee middelste drempels (Step Calibration Measures  $< 1.4$ ) zijn te laag. Item 36 blijkt inconsistenties (misfits) te vertonen. Principale component analyses lieten zien dat enkele items sterkere (cross-)loadings vertoonden op andere factoren zoals Organiseren en Zelfregulatie dan op Elaboreren zelf. Geen van de items, met een gemiddelde van .00 logits (SD = .11 logits) overtreft de gemiddelde persoonsscore (2.29 logits, SD 1.23) hetgeen er op duidt dat de schaal niet het middelste en hogere segment van het continuüm bestrijkt. Het is voor respondenten zeer eenvoudig de items te onderschrijven. Het voorgaande is reden geweest deze schaal verregaand aan te passen. Enkele potentiële items zijn gevonden in de Duitstalige versie van de MSLQ van Wild en Schiefele (1993).

Tabel 12b Elaboreren gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
41	1.25	1.29	.76	.08	.59	Item 39 en 40 zijn afkomstig van Wild en Schiefele (1993) Item 48 is een item van de originele subschaal organiseren
34	.86	.80	.35	.08	.66	
40	1.16	1.27	.30	.08	.60	
37	.76	.70	.20	.08	.69	
39	.95	.87	-.53	.10	.56	
35	.82	.70	-1.09	.10	.63	
All items						
Mean	.97	.94	.00	.09		Person Reliability .62
SD	.18	.25	.62	.01		Person Ability 2.05
All persons						Item Reliability .98
Mean	.96	.94	1.60	.92		Item Difficulty 9.33
SD	1.05	1.08	1.59	.32		Cronbach's alpha .71
Average measures	1	2	3	4	5	
	-1.23	-.57	.38	1.66	4.03	
Step calibration measures		-2.83	-1.10	-.58	4.51	

Met de toevoeging van de items en item 48 uit de schaal Organiseren – die als schaal zelf in het geheel niet voldeed maar waarvan het item 48 hoger laadde op de factor Elaboreren dan op Organiseren – blijkt de schaal iets beter te voldoen aan de Rasch-modelvereisten (zie tabel 12b), maar nog steeds onvoldoende op de Step Calibration Measures. Gezien de enigszins toegenomen Infit en Outfit – die beide prioriteit 1 hebben bij de Raschmodelcriteria – is besloten niet verder te experimenteren met deze schaal en deze als zodanig te handhaven. Het blijft hoe dan ook een van de zwakkere subschalen. Figuur 11 toont de verdeling van personen en items op het onderliggend continuüm.



EACH '#' IS 10.

Figuur 11 Elaboreren gereviseerd: Map van personen en items

### Kritisch denken

De schaal Kritisch denken is samengesteld op basis van drie deelschalen. De schaal is volledig toegesneden op een onderzoeksomgeving. Vier items zijn afkomstig van de MSLQ (Pintrich et al., 1991, 1993), drie items zijn afkomstig van de Duitstalige MSLQ van Wild en Schiefele (1993) en vijf items zijn van eigen hand. Tabel 13a toont de Raschmodelstatistieken van de vier originele MSLQ-items van Pintrich et al. (1991).

Tabel 13a Kritisch denken origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>PTMEA</i>	<i>Inconsistenties</i>	<i>Oordeel</i>
52	.68	.68	.53	.75		
53	.80	.78	.41	.41		dubieus
54r*	.96	.98	-.28	.72	2-1-3-4-5	niet passend
50r*	1.65	1.64	-.65	.56	2-3-1-4-5	niet passend

<i>Kritisch denken - MSLQ</i>						
Person Reliability		.57				
Person Ability		1.85				
Item Reliability		.98				
Item Difficulty		8.40				
Cronbach's alpha		.64				

De items 50 and 54 vertonen inconsistenties in de beantwoording van de categorieën en passen derhalve niet in de context van de schaal. Item 50 heeft ook een Infit en een Outfit > 1.5.

Tabel 13b toont de Raschmodelstatistieken van de drie items van Wild and Schiefele, (1993). De Wild en Schiefele-items voldoen op de Person Reliability Index (< .60) en de Person Ability Scale (< 2) na aan de criteria van het Raschmodel.

Tabel 13b Kritisch denken origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>PTMEA</i>	<i>Inconsistenties</i>
55r	1.14	1.18	.74	.75	-
60	.85	.89	-.13	-.13	-
61	.92	.90	-.61	-.16	-

<i>Kritisch denken - Wild &amp; Schiefele (1993)</i>					
Person Reliability		.56			
Person Ability		1.84			
Item Reliability		.98			
Item Difficulty		9.01			
Cronbach's alpha		.58			



Tabel 13c geeft de vijf zelf ontwikkelde items voor Kritisch denken weer. De items passen volgens alle criteria in het Raschmodel.

Tabel 13c Kritisch denken origineel

Item	Infit	Outfit	Measure	PTMEA
56	.77	.79	.68	.75
58	1.02	1.04	.41	.71
51	1.03	1.02	.35	.73
57	.95	.95	.03	.71
59r	1.23	1.22	-1.46	.58

*Kritisch denken*

Person Reliability	.71
Person Ability	2.43
Item Reliability	.99
Item Difficulty	13.05
Cronbach's alpha	.75

Tabel 13d ten slotte toont alle items tezamen. De items 50 en 54 blijven problemen opleveren. Deze zullen worden verwijderd. Tabel 13e geeft de definitieve schaal weer. De eerder vermelde inconsistenties in item 58 bleken samen te hangen met deze twee verwijderde items.

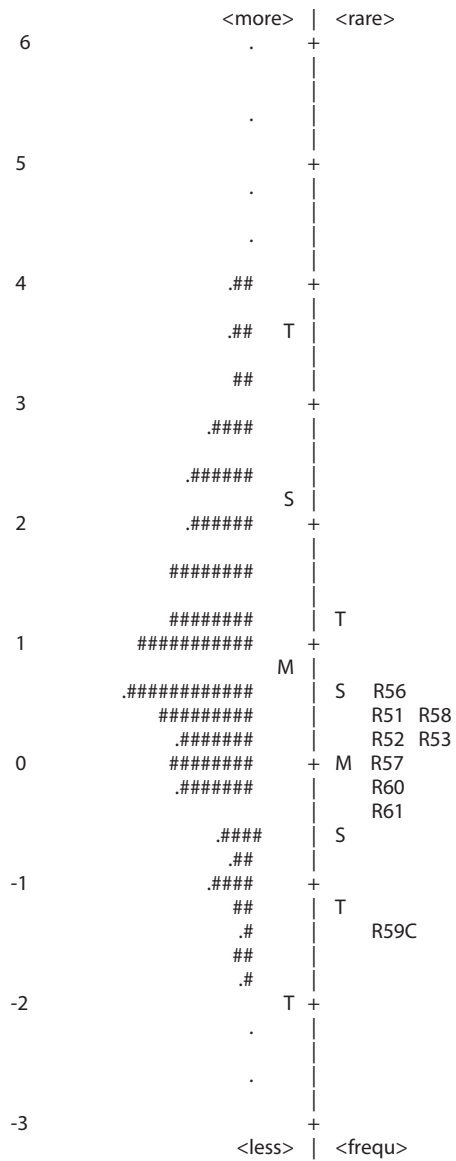
Tabel 13d Kritisch denken origineel

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous
56	.80	.84	.64	.06	.63	58: categorie inconsistenties: 2-1-3-4-5
55r	1.39	1.42	.52	.06	.47	
58	.94	.95	.43	.07	.62	
51	1.01	1.02	.39	.07	.64	54r: inconsistenties: 2-3-1-4-5
52	.87	.89	.27	.07	.61	
53	1.01	1.02	.17	.07	.55	50r: Infit en Outfit liggen buiten de toegestane breedte, tevens categorie inconsistenties: 2-3-4-1-5
57	1.05	1.08	.15	.07	.55	
60	.75	.76	-.03	.07	.65	
61	.62	.61	-.34	.07	.65	
54r	1.04	1.09	-.44	.07	.64	
50r	1.71	1.79	-.76	.07	.43	Person Reliability .80 Person Ability 3.00 Item Reliability .98 Item Difficulty 9.68 Cronbach's alpha .83
59r	.88	.86	-1.00	.07	.60	
All items						
Mean	1.00	1.03	.00	.07		
SD	.28	.30	.51	.01		
All persons						
Mean	1.03	1.03	.82	.48		
SD	.77	.78	1.10	.12		
Average measures	1	2	3	4	5	
	-.37	-.33	.34	1.20	2.55	
Step calibration measures		-2.99	-.84	.43	3.39	

Tabel 13e Kritisch denken definitief

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
56	.84	.87	.66	.07	.68		
58	.98	1.01	.44	.07	.68		
51	1.11	1.11	.38	.07	.67		
52	.99	1.01	.24	.07	.63		
53	1.14	1.13	.12	.07	.58		
57	1.15	1.16	.09	.07	.60		
60	.88	.89	-.12	.07	.66		
61	.71	.70	-.50	.08	.66		
59r	1.18	1.15	-1.30	.08	.56		
All items							
Mean	1.00	1.00	.00	.08		Person Reliability	.79
SD	.15	.15	.56	.01		Person Ability	2.96
All persons						Item Reliability	.98
Mean	1.00	1.00	.81	.60		Item Difficulty	10.04
SD	.80	.80	1.36	.15		Cronbach's alpha	.83
Average measures	1	2	3	4	5		
	1.16	-.73	.25	1.39	3.10		
Step calibration measures		-3.40	-1.18	.52	4.07		

Figuur 13 toont de verdeling van personen en items over het onderliggend continuüm.



EACH '#' IS 4.

Figuur 13 Kritisch denken definitief: map van personen en items

## Organiseren

Tabel 14a geeft de resultaten van de Raschanalyse van de zeven oorspronkelijke items weer.

Tabel 14 Organiseren

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
45	.74	.78	.69	.05	.68		
47	1.11	1.17	.14	.06	.43		
48	.95	.95	.04	.06	.49		
44r	1.49	1.55	.03	.06	.32		
49	.67	.70	-.16	.06	.50		
43	1.36	1.31	-.18	.06	.60		
46	.60	.56	-.57	.07	.59		
All items							
Mean	.99	1.00	.00	.07		Person Reliability	.57
SD	.32	.33	.35	.01		Person Ability	1.87
All persons						Item Reliability	.97
Mean	1.01	1.00	.72	.59		Item Difficulty	7.39
SD	.84	.85	.94	.19		Cronbach's alpha	.60
Average measures	1	2	3	4	5		
	-.66	-.14	.32	.84	1.82		
Step calibration measures		-1.78	-.48	-.49	2.75		

De schaal Organiseren blijkt tamelijk onvoorspelbaar en onbetrouwbaar te zijn gezien de middelste drempelwaarden bij Step Calibration measures en de person reliability index. De person ability index is met 1.87 te laag om twee groepen hoog/laag-scorenden te kunnen onderscheiden. De Rasch maten verklaren slechts 39.8% van de variantie in Organiseren en de step calibration measures zijn gedesorganiseerd. Een Raschmodelanalyse van de schaal na verwijdering van het meest problematische item 44r leverde geen verbetering op. Besloten is geen gebruik meer te maken van deze subschaal.

## Zelfregulatie

Tabel 15a geeft de resultaten van de Raschanalyse van de zeven oorspronkelijke items weer. Deze items zijn afkomstig uit de voorloper van de MSLQ (Pintrich & De Groot, 1990).

Tabel 15a Zelfregulatie origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
63	.90	.93	1.08	.06	.68		
67r	1.07	1.13	.52	.06	.43		
69	1.13	1.16	.29	.06	.49		
64	.97	.97	-.02	.06	.32		
71	.96	.97	-.07	.06	.50		
65	.88	.91	-.71	.07	.60		
70r	1.14	1.07	-1.07	.07	.59		
All items							
Mean	1.01	1.02	.00	.06		Person Reliability	.55
SD	.10	.09	.67	.01		Person Ability	1.81
All persons						Item Reliability	.99
Mean	1.03	1.02	.70	.58		Item Difficulty	14.27
SD	.77	.80	.89	.15		Cronbach's alpha	.57
Average measures	1	2	3	4	5		
	-.68	-.27	.23	.99	2.07		
Step calibration measures		-2.55	-.30	-.17	2.92		

Hoewel de schaal wat betreft de gemiddelde Infit en Outfit goed past, zijn de middelste drempelwaarden in de step calibration measures te weinig onderscheidend en zijn de person reliability index ( $< .65$ ) en de person ability scale ( $< 2$ ) aan de lage kant. De schaal voldoet dus niet aan het minimumcriterium van tenminste twee strata om personen te onderscheiden die hoog en laag scoren op zelfregulatie. Ook Cronbach's alpha - hoewel die geen prioriteit heeft in een Raschmodel - is laag. Een oplossing zou gezocht kunnen worden door verlenging van de schaal met extra items. Zes items zijn afkomstig uit de MSLQ zelf (Pintrich et al., 1991), drie items uit de Duitstalige MSLQ van Wild en Schiefele (1993) en een item - na een SEM analyse waarin bleek dat item 38 meer laadde op Zelfregulatie dan op de oorspronkelijke schaal Elaboreren - van eigen hand. Tabel 15b toont de uitgebreide schaal voor Zelfregulatie.

Tabel 15b Zelfregulatie, uitgebreid

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
63	.95	.99	1.20	.05	.42	66r: category misfits: 2-1-3-4-5
67r	.93	.96	.70	.05	.55	
66r	1.22	1.31	.58	.05	.42	70r: category misfits: 2-1-3-4-5
69	1.22	1.31	.49	.06	.41	
62	1.00	1.09	.38	.06	.37	
64	.98	1.02	.21	.06	.36	
68r	.94	.96	.17	.06	.60	
71	.90	.92	.16	.06	.61	
74	1.03	1.03	.16	.06	.51	
65	.83	.92	-.42	.07	.33	
70r	1.06	1.01	-.75	.07	.56	
38	.53	.54	-.83	.07	.47	
73	1.51	1.46	-1.01	.07	.46	
72	1.17	1.08	-1.05	.08	.56	
All items						
Mean	1.02	1.04	.00	.06		Person Reliability .72
SD	.22	.21	.67	.01		Person Ability 2.48
All persons						Item Reliability .99
Mean	1.05	1.04	.83	.36		Item Difficulty 14.69
SD	.71	.72	.79	.05		Cronbach's alpha .75
Average measures	1	2	3	4	5	
	-.32	-.06	.31	.97	2.01	
Step calibration measures		-2.09	-.26	-.29	2.65	

De uitgebreide schaal leverde echter drie problemen op. Ten eerste waren de step calibration measures niet meer monotoon stijgend. Ten tweede bleek de Infit van (het toegevoegde) item 73 te hoog. Ten derde bleken twee items (66 en 70) inconsistenties in hun oplopende categorieën te vertonen.

Ter oplossing is als eerste (het toegevoegde) item 73 verwijderd. Hiermee werd bereikt dat de step calibration measures weer monotoon stijgend werden. Doordat de inconsistenties in de categorieën van de items 66 (toegevoegd) en 70 (origineel) bleven bestaan, zijn vervolgens ook deze twee items verwijderd. Tabel 15c geeft de resulterende subschaal met Raschmodelstatistieken weer. Deze tabel gaf vanwege zuinigheidsoverwegingen aanleiding ook de items 64 (oorspronkelijk) en 74 (toegevoegd) uit de subschaal te verwijderen. Beide items hebben de laagste PTMEA's in het centrum van de subschaal. Tabel 15d toont de uiteindelijke schaal voor zelfregulatie met negen items.

Tabel 15c Zelfregulatie, tussentijds

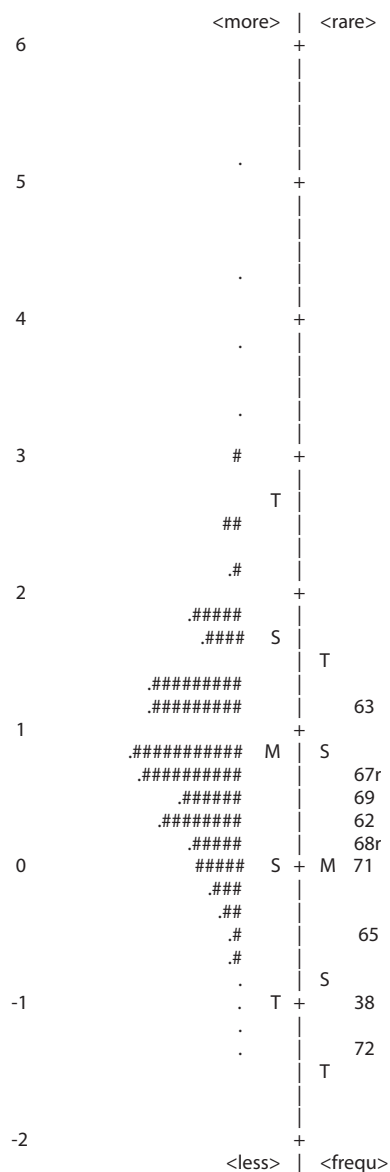
<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
63	.92	.96	1.17	.05	.48	
67r	1.03	1.08	.64	.06	.53	
69	1.21	1.28	.41	.06	.47	
62	.99	1.06	.30	.06	.42	
64	.99	1.02	.12	.06	.40	
68r	1.07	1.09	.08	.06	.56	
71	.95	.96	.06	.06	.63	
74	1.08	1.06	.06	.06	.52	
65	.86	.92	-.56	.07	.37	
38	.59	.59	-1.01	.08	.46	
72	1.30	1.19	-1.26	.08	.54	Person Reliability .68
All items						Person Ability 2.27
Mean	1.00	1.02	.00	.06		Item Reliability .99
SD	.18	.17	.67	.01		Item Difficulty 14.51
All persons						Cronbach's alpha .70
Mean	1.03	1.02	.77	.41		
SD	.70	.71	.83	.06		
Average measures	1	2	3	4	5	
	-.53	-.16	.28	.98	2.03	
Step calibration measures		-2.30	-.31	-.27	2.88	

Tabel 15d Zelfregulatie, definitief

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
63	.91	.95	1.23	.06	.51	
67r	1.03	1.08	.68	.06	.55	
69	1.21	1.27	.44	.06	.49	
62	1.04	1.10	.32	.06	.42	
68r	1.04	1.06	.09	.06	.59	
71	.93	.94	.08	.06	.64	
65	.92	1.00	-.56	.07	.34	
38	.60	.60	-1.02	.08	.46	
72	1.28	1.17	-1.26	.08	.55	
All items						
Mean	1.00	1.02	.00	.06		Person Reliability .64
SD	.18	.18	.76	.01		Person Ability 2.12
All persons						Item Reliability .99
Mean	1.03	1.02	.84	.46		Item Difficulty 15.49
SD	.74	.76	.90	.07		Cronbach's alpha .66
Average measures	1	2	3	4	5	
	-.69	-.24	.27	1.09	2.21	
Step calibration measures		-2.52	-.32	-.21	2.95	

Ondanks alle inspanningen ziet het er niet naar uit de schaal van Zelfregulatie krachtig verbeterd is. Infit en Outfit van items en personen blijven goed en de person reliability index is in waarde toegenomen (hoewel nog steeds te laag: PRI < .65), maar de schaal voldoet nog steeds niet aan alle voorwaarden die het Raschmodel stelt. Wel

is de spreiding van de items over het continuüm beter geworden (van -1.07 tot 1.08 naar -1.26 tot 1.23 logits). Gezien de prioriteitsstelling bij Raschmodelcriteria waarin Infit en Outfit op de eerste plaats komen, kan echter wel besloten worden met deze schaal verder te gaan. Figuur 14 toont de verdeling van de respondenten en items op het continuüm.



EACH '#' IS 5.

Figuur 14 Zelfregulatie gereviseerd: Map van personen en items



*Leeruitkomsten***Autonomie**

Tabel 16a geeft de resultaten van de Raschanalyse van de negen oorspronkelijke items weer.

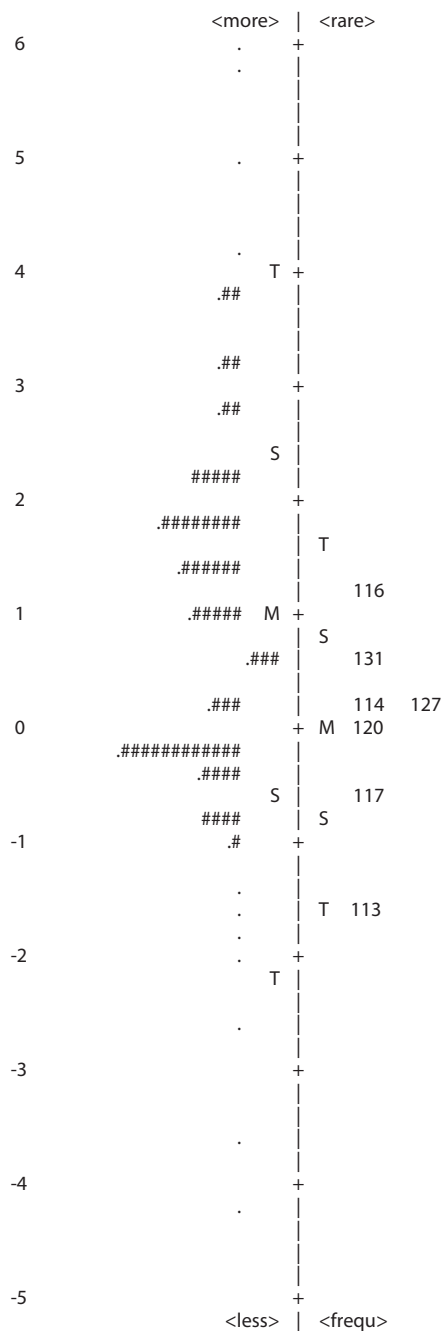
Tabel 16a Autonomie origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
119	1.16	1.21	1.39	.07	.55	
124	1.05	1.09	.88	.07	.48	
116	1.07	1.08	.82	.07	.65	
131	1.23	1.25	.17	.07	.63	
127	.73	.73	-.15	.07	.74	
114	1.09	1.08	-.20	.07	.66	
120	.60	.60	-.37	.07	.72	
117	.86	.86	-.80	.07	.65	
113	1.22	1.16	-1.73	.08	.60	
All items						
Mean	1.00	1.01	.00	.08		Person Reliability .77
SD	.21	.21	.90	.00		Person Ability 2.70
All persons						Item Reliability .99
Mean	1.02	1.01	.66	.59		Item Difficulty 16.19
SD	.93	.91	1.27	.15		Cronbach's alpha .81
Average measures	1	2	3	4	5	
	-1.94	-.75	.07	1.48	2.76	
Step calibration measures		-2.84	-1.51	-.87	3.48	

Op de geringe spreiding in de middelse twee step calibration measures na voldoet de subschaal Autonomie aan de Raschmodelvereisten. Door de items 119 en 124 met de laagste PTMEA's uit de subschaal te verwijderen, worden de step calibration measures enigszins evenwichtiger verspreid, hoewel nog steeds niet met een afstand van minstens 1.4 logits (zie tabel 16b). Figuur 15 toont de verdeling van items en respondenten over het continuüm.

Tabel 16b Autonomie gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
116	1.08	1.13	1.21	.07	.66		
131	1.26	1.30	.55	.07	.64		
127	.74	.77	.20	.08	.74		
114	1.06	1.06	.15	.08	.70		
120	.69	.72	-.03	.08	.69		
117	.91	.92	-.52	.08	.67		
113	1.27	1.19	-1.57	.08	.63		
All items							
Mean	1.00	1.01	.00	.08			
SD	.21	.20	.81	.01			
All persons						Person Reliability	.77
Mean	1.01	1.01	.94	.71		Person Ability	2.77
SD	1.03	1.03	1.54	.19		Item Reliability	.99
Average measures	1	2	3	4	5	Item Difficulty	13.75
	-1.94	-.76	.02	1.67	3.32	Cronbach's alpha	.81
Step calibration measures		-2.60	-1.93	-.74	3.79		



EACH '#' IS 7.

Figuur 15    Autonomie gereviseerd : Map van personen en items

## Afhankelijkheid

Tabel 17a geeft de resultaten van de Raschanalyse van de tien oorspronkelijke items weer.

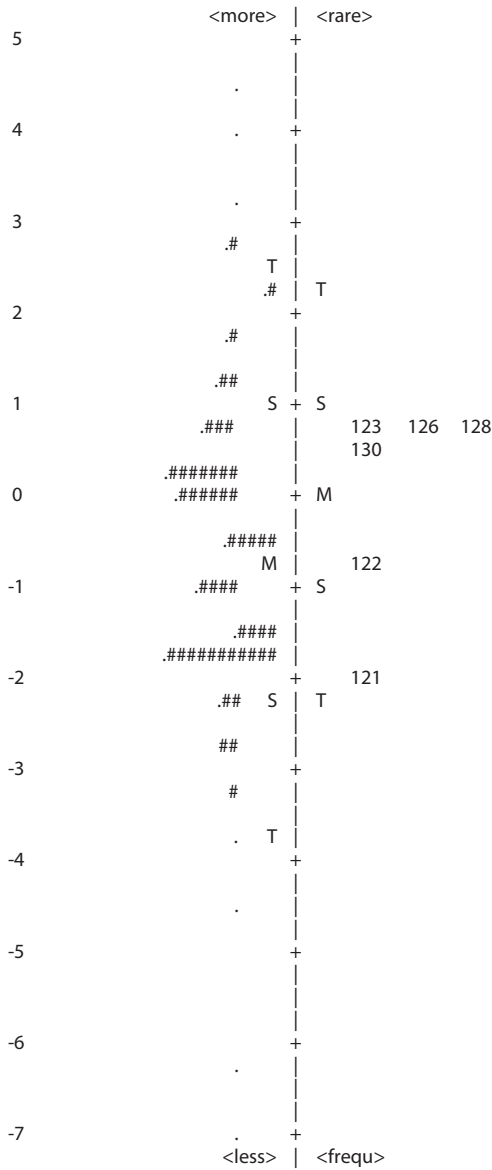
Tabel 17a Afhankelijkheid origineel

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
129	.93	.92	.82	.07	.64	
126	.86	.85	.76	.07	.69	
128	.81	.82	.72	.07	.69	
123	.93	.93	.60	.07	.59	
130	.85	.84	.50	.07	.76	
125	1.35	1.42	.44	.07	.50	
115	1.13	1.19	.42	.07	.63	
122	1.18	1.19	-.64	.07	.55	
121	.85	.85	-1.70	.07	.67	
118	1.02	1.02	-1.92	.07	.61	
All items						
Mean	.99	1.00	.00	.07		Person Reliability .79
SD	.17	.19	.99	.00		Person Ability 2.92
All persons						Item Reliability .99
Mean	1.00	1.00	-.53	.56		Item Difficulty 18.04
SD	.93	.94	1.26	.14		Cronbach's alpha .84
Average measures	1	2	3	4	5	
	-2.52	-1.64	-.27	1.05	1.95	
Step calibration measures		-3.60	-1.11	.91	3.79	

De schaal Afhankelijkheid voldoet aan alle Raschmodelvereisten. Vanuit spaarzaamheidsoverwegingen is echter besloten de items 122, 123 en 125 – deze hebben de laagste PTMEA's – te verwijderen uit de schaal. De resterende schaal wordt weergegeven in tabel 17b. Figuur 16 geeft de verdeling van respondenten en items op het continuüm weer.

Tabel 17b Afhankelijkheid gereviseerd

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
126	.88	.87	.84	.08	.73		
128	.86	.86	.80	.08	.73		
123	1.02	1.03	.65	.08	.63		
130	.87	.86	.55	.08	.80		
122	1.29	1.30	-.80	.08	.59		
121	1.03	1.02	-2.04	.08	.66		
All items							
Mean	.99	.99	.00	.08		Person Reliability	.73
SD	.15	.16	1.07	.00		Person Ability	2.51
All persons						Item Reliability	.99
Mean	1.00	.99	-.63	.78		Item Difficulty	18.23
SD	1.03	1.02	1.54	.19		Cronbach's alpha	.79
Average measures	1	2	3	4	5		
	-3.16	-1.95	-.31	1.36	2.36		
Step calibration measures		-4.14	-1.35	1.15	4.35		



EACH '#' IS 8.

Figuur 16 Afhankelijkheid gereviseerd: map van personen en items

### Statistiekkenistoets

In de vragenlijst zijn negen opdrachten opgenomen die de functionele kennis van de statistiek toetsen. De uitwerkingen van deze opdrachten zijn volgens de SOLO-taxonomie (Biggs, 2003) geïnclassificeerd. Vanwege de vele missing values die deze opgaven kenmerken (item 145 is bijvoorbeeld door 56% van de respondenten niet beantwoord) is ervan afgezien voor deze opgaven het programma Winsteps default op basis van de wel-beantwoorde items een score op vaardigheidsniveau te laten schatten. Niet-beantwoorde opgaven zijn op 0 gesteld en geïnclassificeerd als categorie 'E' (prestructureel, student mist het punt). Tabel 18a geeft de uitkomsten van de Raschanalyses van deze negen opdrachten op de Statistiekkenistoets

Tabel 18a Statistiekkenistoets van de negen oorspronkelijke opdrachten

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous
14	.77	.95	.77	.08	.74	
13	.74	.92	.74	.08	.78	
11	1.11	1.09	.55	.08	.73	
12	1.06	1.02	-.63	.09	.70	
10	.82	.75	-.69	.09	.78	
9	1.15	1.09	-.74	.09	.67	
All items						
Mean	1.00	.97	.00	.08		Person Reliability .80
SD	.12	.12	.69	.01		Person Ability 3.00
All persons						Item Reliability .98
Mean	1.51	.83	1.51	.83		Item Difficulty 11.13
SD	1.91	.22	1.91	.22		Cronbach's alpha .85
Average measures	1	2	3	4	5	
	-2.58	1.00	.20	1.94	4.15	
Step calibration measures		-3.33	-1.15	.02	4.46	

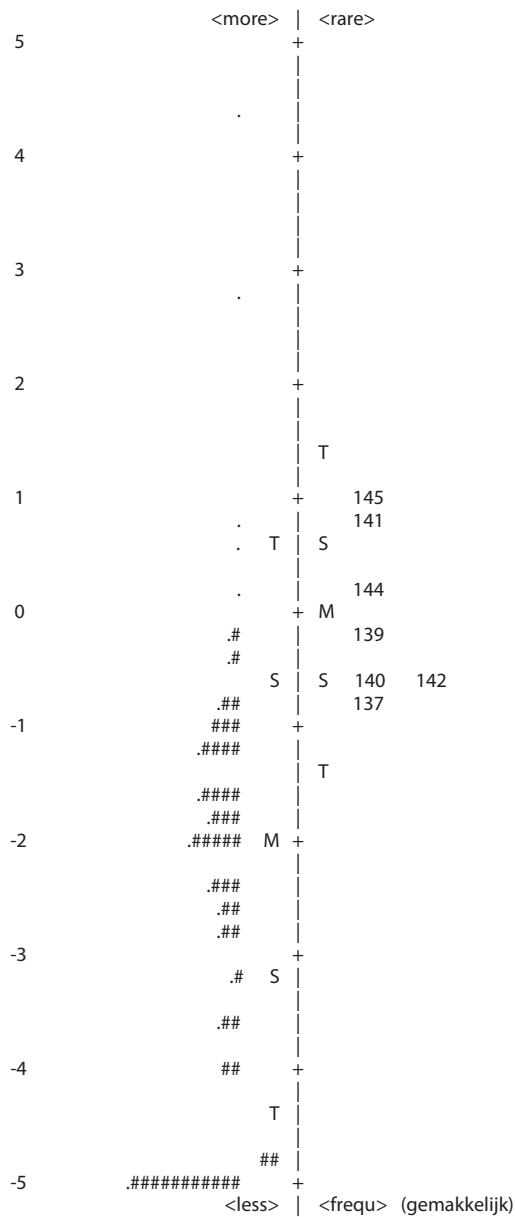
Beide meerkeuze-items (items 143 en 138) blijken in het Raschmodel problematisch te zijn. Ten eerste hebben beide items te hoge Infit en Outfit ( $> 1.5$ ). Ten tweede blijkt item 138 een te hoge outfit voor een van de categorieën te vertonen en laat item 143 inconsistenties zien in de antwoordcategorieën. Op deze items is echter wel het vaakst gescoord: de items bevatten drie respectievelijk elf maal zo veel antwoorden dan de andere zeven opgaven. De bijdrage aan de Statistiekkenistoets is daarmee substantieel: 45% tot 50% van de antwoorden op deze items worden op de Statistiekkenistoets geïnclassificeerd als score "B" ('relationeel niveau'). In de andere zeven opgaven liggen deze percentages tussen 2% en 16%. Deze discrepantie vormt een bedreiging voor de interne consistentie, validiteit en betrouwbaarheid van de statistiekttoets. Beide items dienen derhalve uit de schaal te worden verwijderd. Een consequentie daarvan is dat de test 'moeilijker' wordt, aangezien beide meerkeuzevragen tevens de gemakkelijkste items blijken te zijn (zie de measures in tabel 18a, respectievelijk -1.64 en -1.57 logits).

Na verwijdering van de meerkeuzevragen en recalibratie blijken de zeven resterende opgaven goed te passen in het Raschmodel. De schaal blijkt wel redelijk moeilijk te zijn. Terwijl de items zich op het niveau van 'relationeel' volgens de SOLO-taxonomie bevinden, weerspiegelen de schattingen van de persoonsvaardigheden zich op unistruktuureel en multistruktuureel niveau (zie de verdeling van respondenten en items in figuur 17).

Tabel 18b Statistiekennistoets definitief

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
145	1.24	1.24	1.04	.08	.56		
141	.98	.85	.82	.08	.67		
144	.95	.95	.23	.07	.71		
139	.78	.77	-.15	.07	.79		
142	.87	.90	-.54	.07	.79		
140	.91	.88	-.56	.07	.80		
137	1.31	1.32	-.85	.07	.76		
All items							
Mean	1.00	.99	.00	.08		Person Reliability	.75
SD	.18	.19	.67	.01		Person Ability	2.63
All persons						Item Reliability	.99
Mean	1.00	.99	-1.94	.62		Item Difficulty	11.77
SD	.54	.55	1.26	.14		Cronbach's alpha	.87
Average measures	0	1	2	3	4		
	-3.27	-2.07	-1.17	-.33	2.06		
Step calibration measures		-2.81	-1.58	.45	3.94		





EACH '#' IS 8.

Figuur 17 Map van personen en items voor de Statistiekennistoets

*Samenvatting van de Raschanalyses*

De Raschanalyses hebben ertoe geleid dat van de 16 schalen en 107 items uit de vragenlijst er één schaal (Organiseren) en netto  $36 - 23 = 13$  items niet in het theoretisch model konden worden opgenomen. Tabel 19 geeft de resultaten weer.

Tabel 19 Resultaten van Raschanalyses op van oorsprong Likertitems en -schalen die alle voldoen aan de Raschmodelcriteria. De nummers verwijzen naar de vragenlijst in bijlage II.2.

Construct	Component (schaal)	Items van oorspronkelijke schaal	Verwijderde items	Toegevoegde items	Resteert
Attitude	Affect	146, 147, 156, 159, 160, 166	150, 161		146, 147, 156, 159, 160, 166
	Taakwaardering statistiek	150, 152, 153, 155, 157, 158, 161, 164, 170			152, 153, 155, 157, 158, 164, 170
	Cognitieve competentie	148, 154, 165, 168, 169, 172	148		154, 165, 168, 169, 172
	Moeilijkheid	149, 151, 162, 163, 167, 171, 173	162, 163		149, 151, 167, 171, 173
Motivatie	Intrinsieke waarde	4, 5, 6, 8, 76	4, 5, 6, 8, 76	77, 78, 79	77, 78, 79
	Taakwaardering	9, 10, 11, 12, 13, 14	9, 10, 12	5, 8, 76	5, 6, 11, 13, 14, 76
	Zelfeffectiviteit	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26	18, 20, 24, 25, 28		15, 16, 17, 19, 21, 22, 23, 26
	Testangst	27, 28, 29, 30			27, 29, 30
Leerstrategie	Herhalen	31, 32, 33		75	31, 32, 33, 75
	Elaboreren	34, 35, 36, 37, 38	36, 38	39, 40, 41	34, 35, 37, 39, 40, 41
	Organiseren	43, 44, 45, 46, 47, 48, 49	43, 44, 45, 46, 47, 48, 49		
	Kritisch denken	50, 52, 53, 54		51, 55, 56, 57, 58, 59, 60, 61	50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61
	Zelfregulatie	62, 68, 69, 70, 71	70	38, 63, 65, 67, 72,	38, 62, 63, 65, 67, 68, 69, 71, 72
Leeruitkomst	Functionele kennis van de statistiek	137, 138, 139, 140, 141, 142, 143, 144, 145,	138, 143		137, 139, 140, 141, 142, 144, 145,
	Autonomie	113, 114, 116, 117, 119, 120, 124, 127, 131	124, 131		113, 114, 116, 117, 119, 120, 127
	Afhankelijkheid	115, 118, 121, 122, 123, 125, 126, 128, 129, 130	115, 118, 125, 129		121, 122, 123, 126, 128, 130
Resteert	16 schalen	107 items	- 36 items	+ 23 items	94 items 15 schalen

## Discussie

Op een na ('Organiseren') hebben alle gebruikte Likertschalen de stringente Rasch-analyses overleefd, hoewel soms ingrijpende aanpassingen hebben moeten plaatsvinden. Van de 107 items zijn 13 items uit de diverse meetinstrumenten verwijderd en zijn er 23 items in een andere schaal terechtgekomen. Ondanks het voldoen aan de strenge criteria van het Raschmodel, zijn sommige schalen nog steeds niet optimaal. Dit betreft met name de spreiding van de items over de onderliggende continua. Enkele schalen bevatten items waarmee men het relatief gemakkelijk mee eens kan zijn, zoals Cognitieve competentie, Moeilijkheid, Taakwaardering, Elaboreren, Kritisch denken, Zelfregulatie en Autonomie. Testangst daarentegen bevat items waarmee men het minder snel eens kan zijn en kan daardoor als te moeilijk worden gekenmerkt. Ook de Statistiekennistoets bevat binnen een beperkt gebied van het continuüm te moeilijke items, hoewel objectief gezien de moeilijkheidsgraad van deze items die van een regulier statistiektentamen niet overschrijdt. Op het punt van de spreiding van de items behoeven de diverse schalen nog verbetering. Overigens zou deze lacune in de schalen met de Likertschaalprocedures niet aan het licht zijn gekomen. Met het passen van de items en de schalen volgens het Raschmodel hebben deze ook aan kwaliteit gewonnen. De eendimensionaliteit, de betrouwbaarheid en de constructvaliditeit van de schalen zijn beter gewaarborgd dan met de Likertschaalprocedures mogelijk zijn geweest.

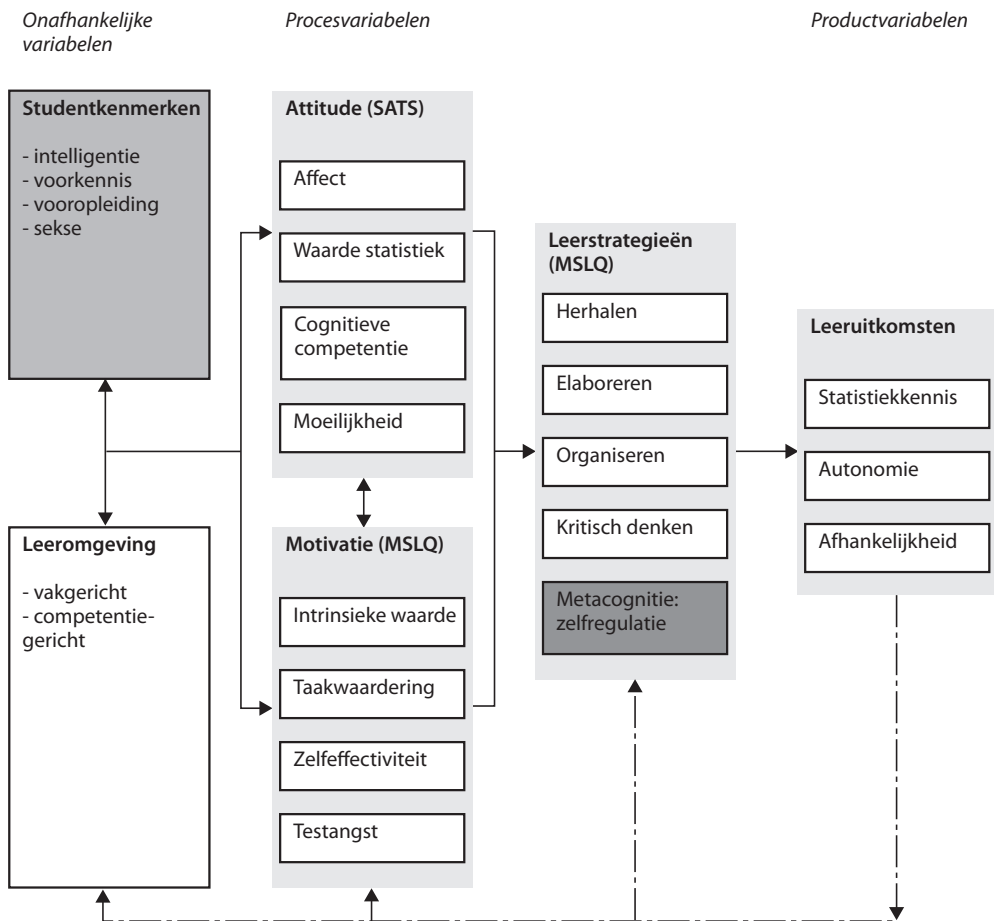
## **Kruisvalidatie van het determinantenmodel met Structural Equation Modeling**

## **Abstract**

Volgens het determinantenmodel (zie figuur 2.1 in dit proefschrift) mediëren attitude, motivatie en leerstrategieën als procesvariabelen de relatie tussen leeromgeving en leeruitkomsten. Deze procesvariabelen worden beschouwd als causale mechanismen waardoor studenten tot bepaalde leeruitkomsten komen. Voor de kruisvalidatie van de causale samenhangen zijn de bestanden van twee groepen die zijn ingedeeld volgens de gevolgde onderwijsbenadering (respectievelijk vakgerichte en competentiegericht statistiekonderwijs) tot één bestand samengevoegd ( $N = 468$ ) en daarna volgens toeval weer opgedeeld in twee even grote steekproeven. Vervolgens is in de calibratiesteekproef het determinantenmodel getoetst dat daarna in de validatiesteekproef is gerepliceerd. Het resulterende structurele model zoals dat uit de kruisvalidering in de validatiesteekproef tevoorschijn is gekomen, zal dienen als uitgangspunt voor de vergelijking van de causale patronen binnen de vakgerichte en de competentiegerichte leeromgeving.

## **Inleiding**

Volgens het determinantenmodel (zie figuur 2.1 in het proefschrift) mediëren attitude, motivatie en leerstrategieën als procesvariabelen de relatie tussen leeromgeving en leeruitkomsten. Deze procesvariabelen worden beschouwd als causale mechanismen waardoor studenten tot bepaalde leeruitkomsten komen. In dit onderzoek worden twee leeromgevingen onderscheiden, te weten de vakgerichte en de competentiegerichte leeromgeving. Voor een beschrijving daarvan wordt verwezen naar paragraaf 5.1. Figuur 1 geeft het determinantenmodel weer met de geoperationaliseerde variabelen. Met meer-groepenanalyse in Structural Equation Modeling (SEM, Byrne, 2001) zullen de effecten van elk van de twee doorlopen leeromgevingen op de procesvariabelen en de leeruitkomsten worden onderzocht. Voor de validatie van het determinantenmodel wordt voorafgaand aan deze exploratie de causale structuur van het model op invariantie onderzocht. Voor deze kruisvalidatie wordt eveneens gebruikt gemaakt van SEM-meer-groepenanalyse waarbij de het totale bestand aan respondenten volgens toeval wordt verdeeld over een calibratie- en een validatiesteekproef (Byrne, 2001; Cudeck & Browne, 1983). Gezien het aantal respondenten in verhouding tot het aantal variabelen in het model wordt afgezien van toetsing van de psychologische constructen en leeruitkomsten als latente en de meetinstrumenten als manifeste variabelen.



Figuur 1 Geoperationaliseerd determinantenmodel

## Methoden

### Participanten

Studenten die zich in het studiejaar 2003-2004 inschreven voor een of beide statistiek-cursussen zijn in het najaar van 2004 benaderd om mee te doen aan het onderzoek ( $N = 1250$ ). Studenten die zich voor het studiejaar 2004-2005 inschreven voor het nieuwe *Onderzoekspracticum kwantitatieve data-analyse* zijn in het najaar van 2005 benaderd om deel te nemen aan het onderzoek ( $N = 769$ ). Bijgevolg zijn er twee groepen studenten te onderscheiden. De eerste groep volgt de traditionele, vakgerichte cursussen statistiek (VAK). De tweede groep heeft het competentiegerichte onderwijs in het eerste onderzoekspracticum doorlopen (COMP). Van de 2019 verzonden vragenlijsten zijn er 468 geretourneerd, een respons van 23%. Van de VAK-studenten zijn dat er 340 (27%) en van de COMP-studenten 128 (16%).

### *Meetinstrumenten*

Voor de operationalisering van de diverse componenten in het determinantenmodel zijn overwegend bestaande meetinstrumenten (Likertschalen) gebruikt. In figuur 5.1 in het proefschrift is het geoperationaliseerde determinantenmodel van figuur 2.1 met de meetinstrumenten weergegeven. Voor een beschrijving van de meetinstrumenten wordt verwezen naar paragraaf 5.2.2. Alle meetinstrumenten voldoen aan de criteria van het Raschmodel (zie bijlage III.2).

### *Procedure*

Voor een beschrijving van de procedure wordt verwezen naar paragraaf 5.2.3 van het proefschrift.

### *Analyse*

Voor de kruisvalidatie van de causale samenhang van het determinantenmodel zijn de bestanden van de twee groepen VAK en COMP eerst tot één bestand samengevoegd en daarna volgens toeval opgedeeld in twee even grote steekproeven. De ene helft fungeert hierbij als de calibratiesteekproef ( $n = 234$ ) en de andere helft als validatiesteekproef ( $n = 234$ ) (Cudeck & Browne, 1983). In de calibratiesteekproef wordt het model getoetst en vervolgens in de validatiesteekproef gerepliceerd. Bij verschillen in het al dan niet significant zijn van paden in calibratie- en validatiesteekproef wordt als uitgangspunt het model gebruikt waarbij alleen paden op nul zijn gezet die in zowel in de calibratie- als in de validatiesteekproef niet significant blijken te zijn. Paden die óf in de calibratie- óf validatiesteekproef significant zijn, worden in de complementaire steekproef, ook als zij daar niet significant bleven, vrij geraamd.

De kruisvalidatie is uitgevoerd met meergroepen-analyse van Structural Equation Modeling (SEM) (Byrne, 2001) met het programma AMOS 5 (Arbuckle, 2003). Als maten voor goedpassen worden de volgende fitindices bij de evaluatie betrokken, te weten: de Goodness-of-Fit Index (GFI), de Comparative Fit Index (CFI) en de RMSEA (Root Mean Square Error of Approximation). De GFI is te interpreteren als het totaal aan verklaarde (co)variantie en zal doorgaans variëren tussen nul en één, waarbij een model beter past naarmate de GFI de waarde één dichter benadert. De CFI (Bentler, 1990) loopt net als de GFI van nul tot één, waarbij een waarde boven de .90 algemeen wordt beschouwd als een indicatie voor een acceptabele fit. De RMSEA (Root Mean Square Error of Approximation) compenseert voor de complexiteit van het model (Browne & Cudeck, 1993). Voor de RMSEA geldt dat waarden onder de .08 worden gezien als een indicatie voor een zeer goede modelpassing. Hu & Bentler (1999) onderzochten voor een groot aantal fitmaten het effect van verschillende grenswaarden. Uit hun onderzoek blijkt dat een gecombineerd criterium van enerzijds een index als de CFI ( $> .95$ ) en anderzijds de RMSEA (maar dan wel  $< .06$ ) tot het kleinste aantal Type I en Type II fouten leidt. Om van een goede fit te mogen spreken moet de CFI de grenswaarde van .95 dicht benaderen en mag de RMSEA niet groter zijn dan .08.

## Resultaten

Tabel 1 geeft de fitindices weer die zijn verkregen met SEM-meer-groepenanalyse. De fitindices betreffen het nulmodel en het geneste model in de calibratie-, validatie- en de totale groep. In alle gevallen is de passendheid van het uiteindelijke model voldoende. Het geneste model zal dienen als uitgangspunt voor de verdere exploratie van het model in VAK en COMP.

Tabel 1 Fitindices voor de passendheid van het theoretisch model in de calibratie-, validatie- en de totale groep

	$\chi^2$	df	$\chi^2/df$	p	GFI	CFI	RMSEA
Calibratie steekproef (n = 234)							
Nulmodel	1379.34	105	13.14	.00	.41		.23
Model	96.06	55	1.75	.00	.95	.97	.06
Validatiesteekproef (n = 234)							
Nulmodel	1536.76	105	14.64	.00	.40	-	.24
Model	87.11	50	1.74	.00	.95	.97	.06
Totale steekproef (N = 468)							
Nulmodel	2916.10	210	13.89	.00	.41	-	.17
Model	235.26	143	1.65	.00	.94	.97	.04

Van de eenendertig significante paden zijn er tweeëntwintig (71%) invariant in zowel de calibratie- als de validatiesteekproef. De  $\chi^2_{(120)}$  van het model met de tweeëntwintig aan elkaar gelijk gestelde padparameters bedraagt 209.52 en de  $\Delta \chi^2_{(40)}$  is 53.09,  $p = .08$ . De negen significant verschillende paden (29%) die in het model vrij zijn geraamd zijn weergegeven in tabel 2.

Tabel 2 Vrij geraamde paden in calibratie- en validatie steekproef

Pad	Calibratie		Validatie		$\Delta \chi^2 (df)$	$p \Delta \chi^2$
	$\beta$	p	$\beta$	p		
Intrinsieke Waarde – Zelfregulatie	.37	.00	.15	.03	$\Delta \chi^2 (4) = 11.47$	.02
Intrinsieke waarde – Statistiekkennis (SOLO)	-.19	.00	naw	ns	$\Delta \chi^2 (24) = 39.40$	.02
Taakwaardering – Elaboreren	.30	.00	.51	.00	$\Delta \chi^2 (6) = 17.88$	.01
Taakwaardering – Kritisch Denken	naw	ns	.22	.00	$\Delta \chi^2 (6) = 13.32$	.04
Taakwaardering – Zelfregulatie	.17	.01	.29	.00	$\Delta \chi^2 (7) = 14.99$	.04
Waarde Statistiek – Herhalen	-.18	.00	naw	ns	$\Delta \chi^2 (15) = 25.17$	.05
Waarde Statistiek – Elaboreren	.18	.01	naw	ns	$\Delta \chi^2 (16) = 27.15$	.04
Moelijkheid – Elaboreren	naw	ns	.23	.00	$\Delta \chi^2 (19) = 29.96$	.05
Zelfregulatie – Afhankelijkheid	naw	ns	.17	.01	$\Delta \chi^2 (39) = 58.43$	.02

Bij de paden Intrinsieke Waarde–Zelfregulatie, Taakwaardering–Elaboreren en Taakwaardering– Zelfregulatie blijken de bèta's van de paden significant te verschillen in de calibratie- en de validatiesteekproef. Bij de overige regressiepaden is het pad



of significant in de calibratiesteekproef, of in de validatiesteekproef, maar is het chi-kwadraat verschil met gelijke parameters significant.

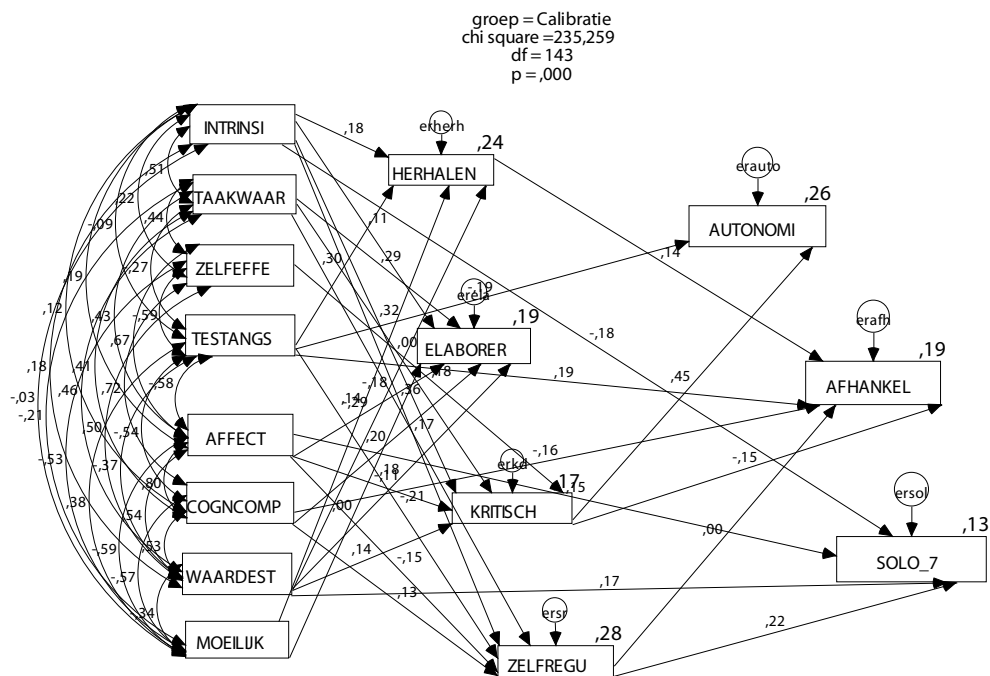
Van de achtentwintig factorcorrelaties zijn er drieëntwintig (82%) invariant in beide steekproeven. De chi-kwadraat<sup>(143)</sup> van het model waarin zowel de tweeëntwintig paden als de drieëntwintig correlaties invariant zijn in zowel de calibratie- als de validatiesteekproef bedraagt 235.26,  $\Delta \chi^2_{(63)} = 78.82$ ,  $p = .08$ . De vijf significant verschillende correlaties (18%) die in het model vrij zijn geraamd zijn weergegeven in Tabel 3.

Tabel 3 Vrij geraamde correlaties in de calibratie- en de validatiesteekproef (2 x n = 234)

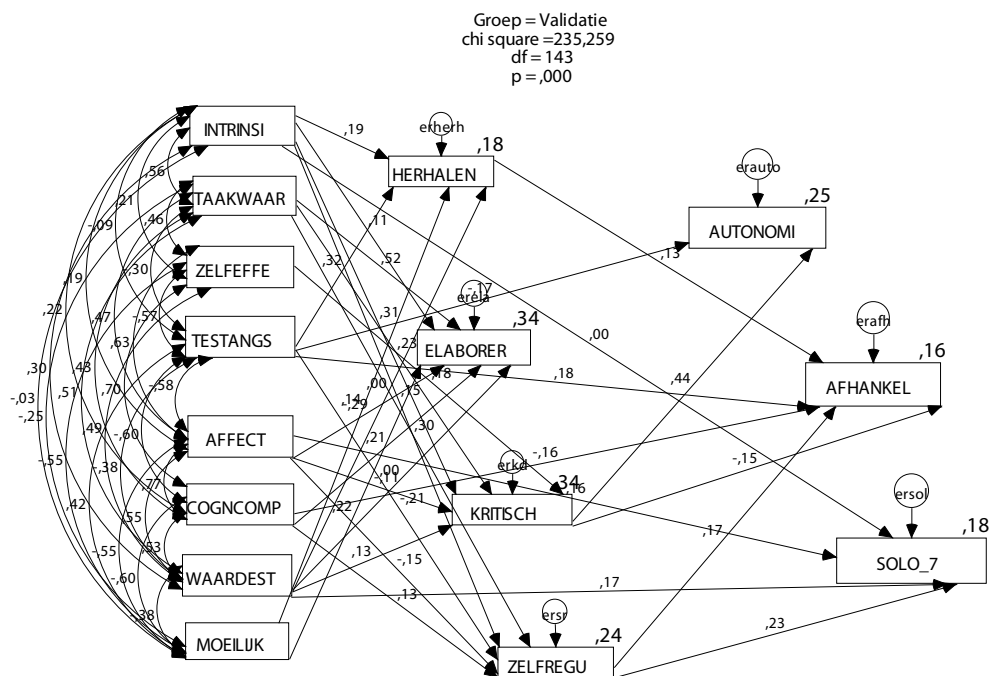
Correlatie	Calibratie		Validatie		$\Delta \chi^2 (df)$	$p \Delta \chi^2$
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>		
Intrinsieke Waarde - Cognitieve Comp.	0.12	.02	0.22	.00	$\Delta \chi^2 (45) = 2.54$	.04
Intrinsieke Waarde - Waarde Statistiek	0.18	.00	0.30	.00	$\Delta \chi^2 (45) = 2.38$	.04
Zelfeffectiviteit – Cognitieve Competentie	0.72	.00	0.70	.00	$\Delta \chi^2 (54) = 3.24$	.04
Testangst – Cognitieve Competentie	-0.54	.00	-0.60	.00	$\Delta \chi^2 (57) = 7.26$	.04
Affect - Moeilijkheid	-0.59	.00	-0.55	.00	$\Delta \chi^2 (61) = 9.67$	.05

Vervolgens zijn paden en correlaties getoetst op invariantie in de calibratie- en validatiesteekproef door de parameters, als de  $\Delta \chi^2$  van het model niet significant bleek op het .05-niveau, aan elkaar gelijk te stellen (Byrne, 2001). Parameters waarbij het significantieniveau van  $\Delta \chi^2$  kleiner of gelijk bleek aan .05, zijn vrij geraamd. Figuur 2 en 3 geven het model weer. In de figuren geven de waarden rechtsboven de rechtehoeken van de procesvariabelen en de leeruitkomsten de proporties verklaarde variantie ( $R^2$ ) weer.

In de tabellen 4a-c en 5a-c zijn de totaaleffecten, de directe effecten en de indirecte effecten van de significante, gestandaardiseerde padcoëfficiënten voor beide steekproeven weergegeven.



Figuur 2 Calibratiesteekproef na invariantietoets



Figuur 3 Validatiesteekproef na invariantietoets

Tabel 4a Calibratiesteekproef: gestandaardiseerde totaaleffecten

	<i>IW</i>	<i>TW</i>	<i>ZE</i>	<i>TA</i>	<i>Aff</i>	<i>CC</i>	<i>Was- ta</i>	<i>Moei</i>	<i>Herh</i>	<i>ELA</i>	<i>KD</i>	<i>Zelf- regu</i>
Herhalen	0,18			0,30			-0,18	0,14				
Elaborer	0,11	0,29			-0,29	0,20	0,18					
Kritisch	0,32		0,18		-0,12		0,14					
Zelfregu	0,36	0,17		-0,21	-0,15	0,13						
Autonomi	0,14		0,08	-0,19	-0,05		0,06				0,46	
Afhankel	-0,02		-0,03	0,23	0,02	-0,16	-0,05	0,02	0,14		-0,15	
Statistiek	-0,10	0,04		-0,05	0,12	0,03	0,17					0,22

Tabel 4b Calibratiesteekproef: gestandaardiseerde directe effecten

	<i>IW</i>	<i>TW</i>	<i>ZE</i>	<i>TA</i>	<i>Aff</i>	<i>CC</i>	<i>Was- ta</i>	<i>Moei</i>	<i>Herh</i>	<i>ELA</i>	<i>KD</i>	<i>Zelf- regu</i>
Herhalen	0,18			0,30			-0,18	0,14				
Elaborer	0,11	0,29			-0,29	0,20	0,18					
Kritisch	0,32		0,18		-0,12		0,14					
Zelfregu	0,36	0,17		-0,21	-0,15	0,13						
Autonomi				-0,19							0,46	
Afhankel				0,19		-0,16			0,14		-0,15	
Statistiek	-0,10				0,15		0,17					0,22

Tabel 4c Calibratiesteekproef: gestandaardiseerde indirecte effecten

	<i>IW</i>	<i>TW</i>	<i>ZE</i>	<i>TA</i>	<i>Aff</i>	<i>CC</i>	<i>Was- ta</i>	<i>Moei</i>	<i>Herh</i>	<i>ELA</i>	<i>KD</i>	<i>Zelf- regu</i>
Herhalen												
Elaborer												
Kritisch												
Zelfregu												
Autonomi	0,14		0,08		-0,05		0,06					
Afhankel	-0,02		-0,03	0,04	0,02		-0,05	0,02				
Statistiek	0,08	0,04		-0,05	-0,03	0,03						

Tabel 5a Validatiesteekproef: gestandaardiseerde totaaleffecten

	<i>IW</i>	<i>TW</i>	<i>ZE</i>	<i>TA</i>	<i>Aff</i>	<i>CC</i>	<i>Was- ta</i>	<i>Moei</i>	<i>Herh</i>	<i>ELA</i>	<i>KD</i>	<i>Zelf- regu</i>
Herhalen	0,19			0,32				0,14				
Elaborer	0,11	0,52			-0,29	0,21		0,22				
Kritisch	0,31	0,23	0,18		-0,11		0,13					
Zelfregu	0,15	0,30		-0,21	-0,15	0,13						
Autonomi	0,13	0,10	0,08	-0,18	-0,05		0,06				0,44	
Afhankel	0,01	0,02	-0,03	0,18	-0,01	-0,14	-0,02	0,02	0,13		-0,15	0,17
Statistiek	0,04	0,07		-0,05	0,13	0,03	0,17					0,23

Tabel 5b Validatiesteekproef: gestandaardiseerde directe effecten

	<i>IW</i>	<i>TW</i>	<i>ZE</i>	<i>TA</i>	<i>Aff</i>	<i>CC</i>	<i>Was- ta</i>	<i>Moei</i>	<i>Herh</i>	<i>ELA</i>	<i>KD</i>	<i>Zelf- regu</i>
Herhalen	0,19							0,14				
Elaborer	0,11	0,52			-0,29	0,21		0,22				
Kritisch	0,31	0,23	0,18		-0,11		0,13					
Zelfregu	0,15	0,30		-0,21	-0,15	0,13						
Autonomi				-0,18							0,44	
Afhankel				0,18		-0,16			0,13		-0,15	0,17
Statistiek					0,16		0,17					0,23

Tabel 5c Validatiesteekproef: gestandaardiseerde directe effecten

	<i>IW</i>	<i>TW</i>	<i>ZE</i>	<i>TA</i>	<i>Aff</i>	<i>CC</i>	<i>Was- ta</i>	<i>Moei</i>	<i>Herh</i>	<i>ELA</i>	<i>KD</i>	<i>Zelf- regu</i>
Herhalen												
Elaborer												
Kritisch												
Zelfregu												
Autonomi	0,13	0,10	0,08		-0,05		0,06					
Afhankel	0,01	0,02	-0,03	0,01	-0,01	-0,02	-0,02	0,02				
Statistiek	0,04	0,07		-0,05	0,04	0,03						

## Discussie

Het determinantenmodel signaleert negen niet-invariante regressiepaden in beide steekproeven, waarvan twee paden ontspringen uit Intrinsieke Waarde, drie vanuit Taakwaardering, twee paden vanuit Waarde van Statistiek, één pad vanaf Moeilijkheid en een laatste pad vanuit Zelfregulatie. Het is niet verontrustend om vast te stellen dat niet-invariante paden voorkomen in de calibratie- en validatiesteekproef, omdat het vermoeden bestaat dat de twee at random samengestelde groepen responden-

ten vanwege de impact van het onderwijsontwerp meer variabiliteit vertonen in de antwoordpatronen dan bij de analyses van VAK en COMP aan het licht zal komen. Een aanwijzing daarvoor is het gezamenlijk model van de calibratie- en de validatiesteekproef. Beide steekproeven blijken bevredigende fit-statistieken te reflecteren (zie tabel 1). In de gezamenlijke groep zijn paden in de calibratiesteekproef versus de validatiesteekproef tweeëntwintig keren gelijk en negen keren verschillend en vijfentwintig blijken in beide steekproeven niet significant en worden voor het ruwe model alvast in beide steekproeven op nul gezet. Dit is al een duidelijke aanwijzing voor grote variabiliteit in beide steekproeven.

De correlaties blijken merendeels (82%) invariant in beide steekproeven en het lijkt erop dat vooral de correlaties tussen motivatiecomponenten en attitudecomponenten een tendens vertonen om zich verschillend te gedragen in de steekproeven.

Het determinantenmodel blijkt talrijke significante paden te vertonen, grotendeels gemedieerd via de leerstrategieën en soms indirect van een motivatiecomponent naar leerresultaten en direct vanuit een attitudecomponent naar leerresultaten. Intrinsieke Waarde heeft als enige motivatiecomponent een rechtstreeks effect op Statistiekkennis (geclassificeerd volgens de SOLO-taxonomie van Biggs en Collis, 1982, zie bijlage III.2) en uitsluitend in de calibratiesteekproef. De negatieve relatie van Intrinsieke Waarde op Statistiekkennis in de calibratiesteekproef geeft te denken, in die zin dat het voor het leerresultaat in competentiegericht onderwijs waarschijnlijk beter is wanneer Intrinsieke Waarde mediërend via de leerstrategieën werkt, en dan vooral via Kritisch Denken, waar het de krachtigste predictor is. Affect ten aanzien van statistiek en Waarde van Statistiek zijn zowel in de calibratiesteekproef als in de validatiesteekproef directe voorspellers van Statistiekkennis. Echter, in beide steekproeven blijkt zelfregulatie een krachtiger voorspeller van het leerresultaat.

Het model met de significant gebleken padcoëfficiënten zal in het proefschrift dienen als uitgangspunt voor de exploratie van de verbanden in VAK en COMP (zie hoofdstuk 5).

## **Voorstudie ijking $T_1$ - $T_2$ -items met Raschmodel**

## Abstract

Bij de studie naar effecten van geïnnoveerd competentiegericht statistiekonderwijs op het leergedrag en de leeruitkomsten van psychologiestudenten zijn op twee momenten in de tijd vragenlijsten gehanteerd die ten dele overeenkomen. Om intra-individuele veranderingen in de tijd te kunnen onderscheiden van veranderingen in de meetomstandigheden is het noodzakelijk dat de meetinstrumenten op beide tijdstippen equivalent aan elkaar zijn. Met de ijkingstrategieën ('anchoring') van het Raschmodel kan worden bewerkstelligd dat niet helemaal overeenkomstige Likertschalen en -items in de twee vragenlijsten toch vergelijkbaar worden gemaakt. Het Raschmodel (voor rating scales zoals Likertschalen is dat de Rasch Rating Scale Model-variant) is een additief lineair model dat de kans beschrijft dat een bepaald persoon een bepaald Likertschaalitem op een specifieke manier zal beantwoorden. De gevolgde procedure is als volgt. Eerst is nagegaan welke (delen van ) schalen en items in beide vragenlijsten inhoudelijk overeenstemmen. Uitgangspunt van deze vergelijking betreft de vragenlijst die in 2002 (tijdstip  $T_1$ ) is afgenomen. Deze vragenlijst bevatte minder schalen en items dan de vragenlijst die in 2005 (tijdstip  $T_2$ ) is voorgelegd aan studenten. Het beperkte aantal items in het  $T_1$ -bestand fungeerde als ondergrens voor de mogelijkheid tot het construeren van de schalen. Wanneer bleek dat het mogelijk was om een beoogde  $T_1$ -schaal met behulp van in de  $T_1$ -vragenlijst aanwezige items te construeren, is dezelfde schaal met dezelfde items geijkt op het bestand van de crosssectionele onderzoek (zie Rasch-voorstudie, bijlage III.2). Ten slotte zijn de schalen van de 193 studenten die in 2005 aan het vervolgonderzoek met dezelfde  $T_2$ -vragenlijst hebben deelgenomen, geijkt op dit crosssectionele bestand.

## Inleiding

Voor de meting van de effecten van geïnnoveerd competentiegericht statistiekonderwijs op het leergedrag en de leeruitkomsten van psychologiestudenten zijn op twee momenten in de tijd vragenlijsten gehanteerd die ten dele overeenkomen. De meetinstrumenten die beide vragenlijsten gemeenschappelijk hebben, gaan over motivatie, leerstrategieën en statistiekkennis. Een probleem voor de vergelijking is dat de vragenlijst van het eerste meetmoment in 2002 (hierna:  $T_1$ -vragenlijst) bij nader beschouwing heeft geleid tot een herziene versie van het instrument dat daarna in 2005 is gebruikt (hierna:  $T_2$ -vragenlijst). Veranderingen ten opzichte van 2002 betreffen onder andere toevoegingen van gevalideerde schalen, zelf ontworpen schalen en items en herformuleringen van items.

In de  $T_2$ -vragenlijst zijn de schalen voor motivatie en leerstrategieën voor het overgrote deel overgenomen van de *Motivated Strategies for Learning Questionnaire* (MSLQ) (Pintrich et al., 1991). De volledige MSLQ bestaat uit een verzameling van gestandaardiseerde Likertschalen met 81 items waarmee motivatie en de daaruit voortkomende leergerichte activiteiten ('leerstrategieën') bepaald worden. Het motivatiedeel beslaat 31 items. Hiermee worden de doelen ('goals') van de studenten, hun waarden ('values') hun overtuigingen van hun vaardigheden om succesvol te zijn in een cursus ('zelfeffectiviteit') en hun testangst bepaald. De vier schalen voor motivatie zijn Intrinsieke

waarde, Taakwaardering, Zelfeffectiviteit en Testangst. Het gedeelte van de MSLQ om leerstrategieën te meten bestaat uit eveneens 31 items die betrekking hebben op de inzet van verschillende cognitieve en metacognitieve strategieën bij het leren. De voor leerstrategieën gebruikte vijf schalen meten Herhalen, Elaboreren, Organiseren, Kritisch denken en Zelfregulatie. In de  $T_1$ -vragenlijst is gebruik gemaakt van de verkorte versie van de MSLQ van Pintrich en De Groot (1990). Deze verkorte versie bestaat uit 44 items die betrekking hebben op een ongedeelde combinatie van Intrinsieke waarde en Taakwaardering, Zelfeffectiviteit, Testangst, Zelfregulatie, een ongedeelde combinatie van Herhalen, Elaboreren en Organiseren (Kritisch denken ontbreekt) en Zelfregulatie. Zie voor een beschrijving van de diverse constructen paragraaf 5.2 van dit proefschrift.

Een groot verschil in de vragenlijsten betreft de (cognitieve) leeruitkomsten. Leeruitkomsten zijn op  $T_2$  gemeten met negen items die beoordeeld en geclassificeerd zijn volgens de *Structure of the Observed Learning Outcomes* (SOLO) taxonomie van Biggs en Collis (1982) (voor een uiteenzetting daarvan zie paragraaf 5.2 van dit proefschrift). De leeruitkomsten op  $T_1$  zijn gemeten met dertien niet op de SOLO-taxonomie gebaseerde items. Deze items beogen voornamelijk te meten of de respondent declaratieve kennis heeft van statistische en methodologische begrippen en methoden. Een voorbeeld is item 1 waarin de respondent dient aan te geven welke onderzoekstermen, zoals verklaarde variantie en frequentieverdeling hem of haar bekend zijn, gescoord met drie antwoordopties (1) 'Ik ken de term wel en weet wat het betekent', (2) 'Ik ken de term, maar weet niet precies wat het betekent' en (3) 'Ik heb nog nooit van deze term gehoord'.

Bij deze verschillen zijn soms schalen zoals die worden verkregen met de op de Klassieke Test Theorie (KTT) gebaseerde Likertschaalprocedures niet meer vergelijkbaar (Roderick & Stone, 1996). Ten tweede, en dat is voor dit onderzoek essentieel, is gebleken dat de meetinstrumenten in de  $T_2$ -vragenlijst na analyse volgens het Raschmodel (Bond & Fox, 2001; Linacre, 2005), drastisch moesten worden gereviseerd voordat ze aan de gehanteerde criteria beantwoordden (zie de Rasch-voorstudie in bijlage III.2).

Samengevat is met de vragenlijsten uit 2002 en 2005 weliswaar gestreefd naar het meten van dezelfde constructen, maar de variabelen uit de  $T_1$ - en  $T_2$ -vragenlijst verschillen in de wijze waarop ze gedefinieerd en geoperationaliseerd zijn. Een noodzakelijke voorwaarde voor het onderzoeken van verschillen tussen studenten op beide meetmomenten is vergelijkbaarheid van betreffende variabelen op die meetmomenten.

De centrale vraag in deze studie is daarom in hoeverre de meetinstrumenten van 2002 te vergelijken zijn met die van 2005. Concreet houdt dat in dat de vraag moet worden beantwoord of er in de  $T_1$ -vragenlijst items te vinden zijn die in voldoende mate hetzelfde latente construct operationaliseren als de schalen uit de  $T_2$ -vragenlijst. Om dit te onderzoeken is in deze studie het Rasch Rating Scale Model (Wright & Masters, 1982) toegepast omdat dit, in tegenstelling tot de analyses uit de KTT, technieken biedt waardoor verschillende versies van een meetinstrument toch kunnen worden vergeleken (Wolfe & Chiu, 1999). Voor een introductie van het Raschmodel wordt verwezen naar de Rasch-voorstudie (bijlage III.2).



## Methoden

### *Participanten*

De  $T_1$ -vragenlijst is in 2002 door 156 studenten ingevuld die statistiek in het competentiegericht onderwijs (IMTO-pilots) volgden en door 666 studenten uit het vakgerichte onderwijs. Deze groepen worden aangeduid met respectievelijk COMP en VAK. Hiervan hebben drie jaar later in 2005 25 studenten uit COMP en 168 uit VAK de  $T_2$ -vragenlijst ingevuld. Voor een beschrijving van deze groep participanten zie paragraaf 6.2.1 in dit proefschrift.

De  $T_2$ -vragenlijst is niet alleen toegezonden aan de studenten die in 2002 de  $T_1$ -vragenlijst hebben ingevuld, maar is ook gebruikt bij studenten die zich in het studiejaar 2003-2004 hebben ingeschreven voor een vakgerichte statistiekcursus of in 2004-2005 voor het competentiegericht onderzoekspracticum *Kwantitatieve data-analyse*. Van COMP hebben er 128 en van VAK 340 studenten de vragenlijsten ingevuld. Voor een beschrijving van deze groep participanten zie paragraaf 5.2.1 in dit proefschrift.

### *Procedure*

Voor de gevolgde procedures zie respectievelijk paragraaf 5.2.2 en paragraaf 6.2.2 van dit proefschrift.

### *Meetinstrumenten*

Voor een overzicht van de gebruikte meetinstrumenten zie paragraaf 5.2 en 6.2 van dit proefschrift.

### *Analyses*

Voor dit onderzoek is gebruik gemaakt van het programma SPSS (Statistical Package for the Social Sciences) versie 12 en voor de Raschanalyses het programma Winsteps (versie 3.6, Linacre, 2005). De Likertschalen zoals deze zijn toegepast in de meetinstrumenten die in deze studie worden onderzocht, vergde in Winsteps het gebruik van het Rasch Rating Scale model (RSM). Het RSM schat persoonsvaardigheden en itemmoeilijkheden voor schalen met meer dan twee ordinaal gerangschikte antwoordcategorieën (Wright & Masters, 1982). In de resultatenparagraaf wordt eerst op inhoudelijke gronden nagegaan welke items uit beide vragenlijsten voldoende overeenstemmen. Daarna wordt met Raschanalyses de unidimensionaliteit, betrouwbaarheid en (construct)validiteit van de schalen beoordeeld. Vervolgens worden de verkregen schalen uit het  $T_1$ -bestand geijkt op dat van het bestand van studenten uit het crossectionele onderzoek (zie de Rasch-voorstudie in bijlage III.2). Ten slotte worden de  $T_2$ -schalen uit het vervolgonderzoek geijkt op dat crossectionele bestand. Op basis van deze geijkte schalen worden in hoofdstuk 6 de verschillen in motivatie, leerstrategieën en leeruitkomsten tussen  $T_1$  en  $T_2$  geanalyseerd en beschreven.

## Resultaten

### Selectie van items

Uitgangspunt voor de selectie van items uit de  $T_1$ -vragenlijst zijn de Rasch-schalen uit bijlage III.2. Daarvan is aangetoond dat ze in voldoende mate unidimensioneel zijn en is de betrouwbaarheid en constructvaliditeit onderbouwd. Tabel 1 bevat een overzicht van de geselecteerde items die zowel in de  $T_1$ - als in de  $T_2$ -vragenlijst voldoende overeenstemmen.

Tabel 1 Overlappende items voor  $T_1$ - $T_2$ -vergelijking. De nummers verwijzen naar de desbetreffende vraag in de vragenlijst.

Itemnr $T_2$	Itemnr $T_1$	Item
<b>Taakwaarde</b>		
8	56	Ik denk dat er interessante dingen te leren zijn in deze modules.
9	46	
		Ik denk dat wat ik leer van statistiek door onderzoek te doen, ook kan gebruiken bij andere vakken.
10	43	Ik vind het belangrijk om zowel onderzoek te leren uitvoeren als statistiek toe te leren passen.
12	54	Ik denk dat wat ik allemaal door het uitvoeren van onderzoek leer, belangrijk is voor de rest van mijn studie.
13	44	Ik leer graag iets over onderzoeksmethoden en statistiek.
14	59	Begrip van statistische procedures is belangrijk voor mij.
<b>Zelfeffectiviteit</b>		
15	52	Ik denk dat hoge cijfers voor de eindopdrachten van deze modules haalbaar zijn.
19	50	Ik ben er zeker van dat ik de onderzoekstaken heel goed zal kunnen uitvoeren.
20	47	Ik verwacht dat ik deze modules met heel goede resultaten zal afronden.
21	2	Vergeleken met andere studenten die onderzoek doen en daarbij statistiek toepassen verwacht ik het beter te zullen doen.
22	45	Ik ben er zeker van dat ik deze modules goed zal kunnen invullen.
23	48	Ik denk dat ik een goede onderzoeker ben in vergelijking met andere studenten die onderzoek moeten doen.
26	57	Ik denk dat ik in staat ben om de theorie in deze modules te begrijpen en daardoor onderzoek uit te kunnen voeren.
<b>Testangst</b>		
27	60	Als ik mijn eindopdracht maak denk ik eraan hoe slecht ik het doe.
28	51	Ik heb een ongemakkelijk en ellendig gevoel als ik een eindopdracht moet maken.
29	3	Ik ben zo gespannen tijdens het maken van mijn eindopdracht, dat ik de dingen die ik geleerd heb niet
30	58	Ik maak me zorgen over de eindopdrachten.

<i>Itemnr</i> <i>T<sub>2</sub></i>	<i>Itemnr</i> <i>T<sub>1</sub></i>	<i>Item</i>
		<b>Leerstrategieën</b>
33	69	Om mezelf te helpen de theoretische informatie te onthouden lees ik mijn aantekeningen telkens over.
34	61	Als ik de theorie en onderzoekstaken bestudeer probeer ik dat met informatie uit andere bronnen te integreren.
35	79	Tijdens het bestuderen van de theorie probeer ik de dingen die ik lees, te verbinden met wat ik al weet.
37	75	Als ik een onderzoekstaak bestudeer, probeer ik dat met informatie uit verschillende bronnen te verbinden.
38	74	Bij het uitvoeren van nieuwe onderzoekstaken maak ik gebruik van wat ik heb geleerd in andere opdrachten.
47	62	Bij het bestuderen van het theoretische gedeelte probeer ik me te herinneren wat op de bijeenkomsten is gezegd, zodat ik de betekenis en achtergronden van het onderzoek zo goed mogelijk begrijp.
49	68	Als ik een eindopdracht voorbereid probeer ik mezelf zo veel mogelijk dingen uit het theoretisch gedeelte te herinneren.
		<b>Zelfregulatie</b>
62	63	Ik stel mezelf vragen om er zeker van te zijn dat ik de theorie die ik heb bestudeerd, beheers.
63	70	Als een onderzoekstaak niet verplicht is, dan maak ik hem toch.
64	73	Voordat ik met de onderzoekstaak begin, denk ik na over de theorieën die ermee verbonden zijn zodat ik de onderzoeksstappen kan bepalen.
65	76	Als ik bij het uitvoeren van onderzoek niet alles begrijp dan maak ik even pas op de plaats om de theorie eromheen nog eens door te nemen.
69	78	Ik werk hard om een hoog cijfer te halen, zelfs al vind ik een onderzoeksopdracht en de theorie niet leuk.
71	71	Ik werk door totdat ik klaar ben, zelfs al zijn de taken saai en vervelend.
		<b>Functionele kennis van de statistiek (SOLO)</b>
140	12	Welke relatie bestaat er volgens u tussen de statistische grootheden 'gemiddelde', 'standaardafwijking' (ook wel standaarddeviatie genoemd) en 'normaalverdeling'? Beschrijf een bekende vuistregel in eigen woorden.

### *Raschanalyses van de schalen*

Tabel 2 bevat een samenvatting van de te hanteren criteria voor de beoordeling van de fit van de items en de schalen aan het Raschmodel. Voor een beschrijving van de criteria wordt verwezen naar de Rasch-voorstudie (bijlage III.2). De criteria dienen simultaan te worden toegepast, waarbij voor de beoordeling van items en schalen in 'twijfelgevallen' of 'grensgebieden' de Infit- en Outfit-criteria het zwaarst wegen, gevolgd door de categorie-indices en de diverse betrouwbaarheden en als laatste de persoonsvaardigheden (Persons Ability Index) en de Itemmoeilijkheden (Item Difficulty Index).

Tabel 2 Criteria voor de beoordeling van voor het Raschmodel acceptabele Likertitems en Likertschalen

<i>Criterion</i>	<i>Acceptabel indien:</i>	<i>Prioriteit</i>
Infit Mean Square	tussen 0.5 – 1.5	1
Outfit mean Square	tussen 0.5 – 1.5	1
Average Measure	monotoon stijgend	2
Step Calibration Measure	monotoon stijgend én > 1.4 logits verschil tussen de drempels	2
Person Reliability Index	≥ .60	2
Person Ability Scale	≥ 2	3
Item Reliability Index	≥ .95	2
Item Difficulty Scale	≥ 2	3
Cronbach's alpha	≥ .70	n.v.t.

*Taakwaardering*

In tabel 3 zijn de resultaten van de Raschanalyse vermeldt. De itemnummers verwijzen naar de vragenlijst uit 2002 (bijlage II.3). Een r achter een itemnummer geeft aan dat de waarden van het oorspronkelijk negatief geformuleerde item zijn gehercodeerd ('reversed'). De itemnummers rechts in de tabel zijn de overeenkomende items in de  $T_2$ -vragenlijst (bijlage II.2).

Tabel 3 Taakwaardering  $T_1$ 

<i>Item <math>T_1</math></i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Item <math>T_2</math></i>	<i>Miscellaneous</i>
59	.90	.93	.64	.06	.80	14	
44	1.24	1.26	.64	.06	.78	13	
56	.97	1.01	.34	.06	.77	8	
43	.89	.88	-.23	.07	.81	10	
54	.86	.83	-.57	.07	.79	12	
46	1.06	1.06	-.82	.07	.75	9	
All items							
Mean	.99	.99	.00	.07		Person Reliability	.83
SD	.13	.14	.57	.00		Person Ability	3.33
All persons						Item Reliability	.99
Mean	.99	.99	1.37	.80		Item Difficulty	11.51
SD	.93	.96	2.05	.24		Cronbach's alpha	.90
Average measures	1	2	3	4	5		
	-2.38	-1.15	.16	1.72	3.97		
Step calibration measures		-2.71	-1.01	-.13	3.86		

De data passen goed in het Raschmodel, met gemiddelde algemene Infit- en Outfit-waarden van .99 en .99 voor zowel de items als de personen. De gemiddelde maten (average measures) en step calibrations zijn in oplopende volgorde. De schaal heeft een itemmoeilijkheid en itembetrouwbaarheid van 11.51 en .99 en een persoonsvaardigheid en een persoonsbetrouwbaarheid van 3.33 en .83. De schaal identificeert derhalve 11.51 verschillende itemmoeilijkheidsstrata, die onderscheiden worden door de personen en 3.33 verschillende persoonsvaardigheidsstrata, die door de

items worden onderscheiden. Beide zijn ruim boven het vereiste van twee strata, dat minimaal nodig is om een zinvol onderscheid te kunnen maken tussen gemakkelijke of moeilijke items of tussen op het gebied van op taakwaardering bekwame of minder bekwame personen. De range van itemcalibraties strekt zich uit van  $-.82$  (item 46) tot  $.64$  (item 59) logits. De geordende oplopende rangschikking van de items levert ondersteuning voor de constructvaliditeit van de schaal, omdat de items logisch zijn gerangschikt van gemakkelijk tot moeilijk.

De gemiddelde item(moeilijkheids)maat is  $.00$  (SD  $.57$ ) en de gemiddelde persoons (vaardigheids)maat is  $1.37$  (SD  $2.05$ ). Het moeilijkste Taakwaardering-item, item 59 ( $.64$  logits) wordt ruim tweemaal overstegen door de gemiddelde persoonsmaat, wat aangeeft dat de schaal erg gemakkelijk te onderschrijven is en het hogere uiteinde in het continuüm van de onderliggende variabele ontoereikend meet. Echter, voor het beperkte aantal items dat in deze studie ter beschikking staat en gezien het feit dat de schaal ook een overeenkomstige goede fit heeft in de  $T_2$  is Taakwaardering een valide construct voor de onderliggende mengvorm van extrinsieke en intrinsieke redenen om betrokken te zijn bij een leertaak.

### *Zelfeffectiviteit*

Bij de eerste exploratie met de origineel in de  $T_1$  aanwezige acht items bleek item 55 ( $T_2$ -tegenhanger item 24 'Mijn onderzoeksvaardigheden zijn uitstekend vergeleken met die van andere studenten die onderzoek uitvoeren en daarbij statistiek toepassen') in de context van de overige items een misfittend item, met een infit van  $1.38$  maar een outfit van  $1.83$ . De Rasch-modellering werd daarom voortgezet met een 7-items construct. Dit items construct bleek zeer goed te passen in het Raschmodel (Tabel 3).

Tabel 3      Zelfeffectiviteit  $T_1$

Item $T_1$	Infit	Outfit	Measure	Error	PTMEA	Item $T_2$	Miscellaneous
52	.86	.82	1.38	.07	.84	15	
48	.96	.98	1.11	.07	.83	23	
2	1.08	1.18	1.02	.07	.80	21	
50	.75	.73	.23	.07	.87	19	
47	.98	.92	.16	.07	.85	20	
45	1.16	1.12	-1.24	.07	.84	22	
57	1.22	1.27	-2.66	.07	.75	26	
All items							
Mean	1.00	1.00	.00	.07		Person Reliability	.91
SD	.15	.18	1.35	.01		Person Ability	4.54
All persons						Item Reliability	1.00
Mean	.97	1.00	-.11	.74		Item Difficulty	25.83
SD	.92	1.04	2.50	.19		Cronbach's alpha	.93
Average measures	1	2	3	4	5		
	-4.18	-2.21	-.38	2.04	4.50		
Step calibration measures		-3.35	-2.06	1.00	4.41		

De data passen zeer goed in het Raschmodel, met gemiddelde algemene Infit- en Outfit statistieken van 1.00 en 1.00 voor de items en .97 en 1.00 voor de personen. De gemiddelde maten (average measures) en step calibrations blijken in oplopende volgorde. De schaal heeft een itemmoeilijkheid en itembetrouwbaarheid van 25.83 en 1.00, en een persoonsvaardigheid en persoonsbetrouwbaarheid van 4.54 en .91. De schaal identificeert 25.83 verschillende itemmoeilijkheidsstrata, die onderscheiden worden door de personen en 4.54 verschillende persoonsvaardigheidsstrata, die door de items worden onderscheiden. Beide zijn ruim boven het vereiste van twee strata, dat minimaal nodig is om een zinvol onderscheid te kunnen maken tussen gemakkelijke of moeilijke items of tussen op het gebied van Zelfeffectiviteit bekwame of minder bekwame personen. De range van item calibraties strekt zich uit van -2.66 (item 57) tot 1.38 (item 52) logits. De geordende oplopende rangschikking van de items onderschrijft de constructvaliditeit van de schaal, omdat de items logisch zijn gerangschikt van gemakkelijk tot heel moeilijk.

De gemiddelde item(moeilijkheids)maat is .00 (SD 1.35) en de gemiddelde persoons (vaardigheids)maat is -.11 (SD 2.50). Het moeilijkste Zelfeffectiviteitsitem is item 52, met 1.38 logits. De gemiddelde itemmoeilijkheid en gemiddelde persoonsmaat liggen redelijk op eenzelfde niveau, waardoor duidelijk wordt dat de schaal zowel het bovenste als onderste segment van het continuüm in Zelfeffectiviteit adequaat meet. De schaal heeft een overeenkomstige goede fit in de  $T_2$ .

### *Testangst*

Voor de operationalisatie van Testangst is uitgegaan van de originele vier items, zoals deze zowel voorkwamen in de  $T_1$ - als in de  $T_2$ -schaal. Bij de eerste exploratie met deze compositie bleek dat testangst op  $T_1$  so wie so minder betrouwbaar ingevuld was dan op  $T_2$ , omdat zelfs bij een vier-itemsschaal de persoonsbetrouwbaarheid op  $T_2$  niet hoger uitkwam dan .68 (de drie-itemschaal op  $T_2$  bereikt een betrouwbaarheid van .71). Overgaan tot een drie-itemschaal gaf geen verbetering, omdat de persoonsbetrouwbaarheid op  $T_1$  dan zelfs afnam tot .57. Voor het zo accuraat mogelijk vergelijken van Testangst in beide bestanden is vanwege de gebrekkige persoonsbetrouwbaarheid op  $T_1$  dan ook gekozen voor de operationalisatie van Testangst met vier items (Tabel 4).

Tabel 4 Testangst  $T_1$ 

Item $T_1$	Infit	Outfit	Measure	Error	PTMEA	Item $T_2$	Miscellaneous
60	1.10	1.03	.64	.06	.74	27	
3	1.22	1.27	.25	.06	.74	29	
51	.78	.75	-.07	.06	.83	28	
58	.95	.92	-.83	.05	.85	30	
All items							
Mean	1.01	.99	.00	.06		Person Reliability	.68
SD	.16	.19	.54	.00		Person Ability	2.25
All persons						Item Reliability	.99
Mean	.99	.99	-.77	.77		Item Difficulty	12.40
SD	.92	.94	1.40	.22		Cronbach's alpha	.84
Average measures	1	2	3	4	5		
	-2.21	-1.21	-.31	.56	1.64		
Step calibration measures		-2.06	-.16	.20	2.02		

De data passen goed in het Raschmodel, met gemiddelde algemene Infit- en Outfit statistieken van 1.01 en .99 voor de items en .99 en .99 voor de personen. De gemiddelde maten (average measures) en step calibrations zijn in oplopende volgorde. De schaal heeft een itemmoeilijkheid en itembetrouwbaarheid van 12.40 en .99, en een persoonsvaardigheid en persoonsbetrouwbaarheid van 2.25 en .68. De schaal identificeert 12.40 verschillende itemmoeilijkheidsstrata, die onderscheiden worden door de personen en 2.25 verschillende persoonsbekwaamheidsstrata, die door de items worden onderscheiden. Beide zijn boven het vereiste van twee strata, dat minimaal nodig is om een zinvol onderscheid te kunnen maken tussen gemakkelijke of moeilijke items of tussen op het gebied van Testangst zeer uitgesproken of minder uitgesproken personen. De range van itemcalibraties strekt zich uit van -.83 (item 58) tot .64 (item 60) logits. De geordende oplopende rangschikking van de items is een indicatie voor de constructvaliditeit van de schaal. De items zijn logisch gerangschikt van gemakkelijk tot moeilijk.

De gemiddelde item(moeilijkheids)maat is .00 (SD .54) en de gemiddelde persoon(s)vaardigheid)maat is -.77 (SD 1.40). Het moeilijkste Testangst item is item 60, met .64 logits. De gemiddelde persoonsmaat ligt ruim onder de gemiddelde itemmoeilijkheid. Dit betekent dat de schaal de onderkant van het continuüm in Testangst onvoldoende goed meet. Zoals gezegd is de persoonsbetrouwbaarheid van deze schaal op  $T_2$  hoger.

#### *Leerstrategieën*

Op  $T_1$  werd nog geen onderscheid gemaakt tussen de verschillende leerstrategieën. De  $T_1$ -vragenlijst was een wat minder ideaal vertaald exemplaar van de verkorte MSLQ van Pintrich & De Groot (1990). Gezien het beperkte aantal items dat op  $T_1$  beschikbaar was voor het combinatieconstruct voor leerstrategieën had het geen zin om aparte schalen te maken, vooral omdat de items uit het  $T_1$ -combinatieconstruct slechts 2 herhaalitems bevatte, naast 5 items voor elaboreren en organiseren elk. Kri-

tisch denken maakte geen deel van het construct. Uit de Rasch-modeling op  $T_2$  bleek dat juist de items uit elaboreren en organiseren de zwakste items waren. Om toch een serieuze poging te ondernemen om dan maar één algemeen Leerstrategieën-construct te maken, zijn in eerste instantie, net als Pintrich en De Groot (1990), de twaalf items alle onder in één operationalisatie gebracht.

Zoals verwacht rezen er problemen bij het construct. Het enige omgepoolde item in het construct, item 64 'Ik vind het moeilijk om de hoofdpunten te ontdekken in de theorie' (een voormalig organisatie-item) bleek een infit 1.57 en outfit 1.97 te hebben, terwijl item 77 'Als hulp bij het bestuderen van het theoretisch gedeelte maak ik een samenvatting van de kernpunten daaruit' eveneens een infit 1.53 en outfit 1.49 bleek te hebben. Er kwamen tevens allerlei misfits in de schaal voor en de 'category thresholds' bleken voortdurend te overlappen. Daarom zijn er een aantal items verwijderd. Dit patroon komt overeen met het modelleren in Rasch van de Elaboratie en Organisatie-items (zie de Rasch-voorstudie, bijlage III.2). De items bleken voortdurend misfits te vertonen, totdat vijf items volledig waren verwijderd. Daarna waren geen misfits meer en bleven de in- en outfits binnen de norm. Zelfs de 'thresholds' waren in oplopende volgorde. Leerstrategieën op  $T_1$  bleek een relatief zwak construct, en dat vertoonde een parallel met de op  $T_2$  geconstrueerde operationalisatie van Elaboreren. Echter, op  $T_1$  moest volstaan worden met deze items. Het uiteindelijke  $T_1$ -construct ziet er nu als volgt uit (Tabel 5).

Tabel 5 Leerstrategieën  $T_1$

Item $T_1$	Infit	Outfit	Measure	Error	PTMEA	Item $T_2$	Miscellaneous
69	1.30	1.38	.57	.05	.55	33 Herhalen	
75	.80	.88	.51	.05	.61	37 Elaboreren	
61	.96	.99	.31	.05	.58	34 Elaboreren	
62	1.04	1.14	.05	.06	.54	47 Organiseren	
68	1.01	1.01	.00	.06	.58	49 Organiseren	
74	.90	.89	-.55	.07	.53	38 Elaboreren	
79	.86	.81	-.89	.07	.55	35 Elaboreren	
All items							
Mean	.98	1.01	.00	.06		Person Reliability	.61
SD	.15	.18	.50	.01		Person Ability	1.99
All persons						Item Reliability	.99
Mean	1.03	1.01	1.46	.76		Item Difficulty	11.29
SD	.88	.84	1.12	.22		Cronbach's alpha	.72
Average measures	1	2	3	4	5		
	-.28	.18	.59	1.25	2.47		
Step calibration measures		-1.29	-.59	-.51	2.32		

De data passen nu redelijk in het Raschmodel, met gemiddelde algemene Infit- en Outfit statistieken van .98 en 1.01 voor de items en 1.03 en 1.01 voor de personen. De gemiddelde maten (average measures) en step calibrations blijken in oplopende volgorde. De schaal heeft een itemmoeilijkheid en itembetrouwbaarheid van 11.29 en .99, en een persoonsvaardigheid en persoonsbetrouwbaarheid van 1.99 en .61. De schaal



identificeert 11.29 verschillende itemmoeilijkheidsstrata, die onderscheiden worden door de personen en 1.99 verschillende persoonsbekwaamheidsstrata, die door de items worden onderscheiden. De itemmoeilijkheidsstrata zijn ruim boven het vereiste van twee strata, dat minimaal nodig is om een zinvol onderscheid te kunnen maken tussen gemakkelijke of moeilijke items. De persoonsbekwaamheidsstrata zijn net aan toereikend voor het maken van een zinvol onderscheid tussen op het gebied van leerstrategieën bekwame of minder bekwame personen. De range van item calibraties strekt zich uit van -.89 (item 79) tot .57 (item 69) logits. De geordende oplopende rangschikking van de items is een indicatie voor enige constructvaliditeit van de schaal, omdat de items logisch zijn gerangschikt van gemakkelijk tot heel moeilijk. De gemiddelde item(moeilijkheids)maat is .00 (SD .50) en de gemiddelde persoon(vaardigheids)maat is 1.46 (SD 1.12). Het moeilijkste Leerstrategieën-item is item 69, met .57 logits. De gemiddelde itemmoeilijkheid en gemiddelde persoonsmaat liggen ver van elkaar verwijderd en de gemiddelde persoonsmaat overstijgt het moeilijkste item bijna drie maal. Dit maakt duidelijk dat de schaal veel te gemakkelijk te onderschrijven is en het bovenste deel van het continuüm in leerstrategieën inadequaat meet. De schaal heeft een overeenkomstige fit op  $T_2$ .

### *Zelfregulatie*

Zelfregulatie is op  $T_1$  een construct dat op het randje ligt van acceptabiliteit in het Raschmodel. Voor deze operationalisatie waren er op  $T_1$  slechts zeven items beschikbaar. De eerste operationalisatie in het Raschmodel met die originele zeven items had een persoonsbetrouwbaarheid van .58 en een persoonvaardigheid van 1.88. Het enige omgepoolde item echter had een disharmonieuze infit van 1.49 en een outfit van 1.42. Wanneer dit item werd verwijderd, daalde de persoonsbetrouwbaarheid met slechts 1/100 tot .57, maar waren alle in- en outfits harmonieuzer. Bij Rasch ligt er meer accent op de fit van de In- en Outfits, dan op het betrouwbaarheidsconcept zoals in de Klassieke Testtheorie (Linacre, 2005). Op basis daarvan werd ervoor gekozen om bij Zelfregulatie op  $T_1$  de toetsing voort te zetten met zes items met harmonieuze in- en outfits, ondanks de fractioneel zwakkere persoonsbetrouwbaarheid. De thresholds bleken echter onregelmatig. Categorie 2 bleek nimmer de meest voorkomende categorie. Om het probleem met overlappende thresholds op te lossen en een passende schaal met een acceptabele fit te verkrijgen, mogen probleemcategorieën bijeengevoegd worden in daarop volgende categorieën, omdat de overmaat aan categorieën geen correcte meting blijken voor de onderliggende variabele (Linacre, 2005). Bij de Zelfregulatieschaal op  $T_1$  betekende dit dat de voormalige Likertcategorieën 1-5 omgevormd zijn tot drie categorieën, die zich uitstrekken van 3 tot 5, waarbij de categorieën 1 en 2, waarin nauwelijks scores voorkwamen, zijn opgegaan in één categorie 3. De schaal met drie categorieën paste in het Raschmodel. De aanpassing van de categorieën verlaagde de (te) lage persoonsbetrouwbaarheid naar .54, maar de alfa (wat al een ondergrens is voor persoonsbetrouwbaarheid) was nog wel voldoende en de in- en outfits bleken harmonieus, terwijl de thresholds nu ook monotoon verliepen. Uiteindelijk bleek deze optie de minst problematische oplossing (Tabel 6).

Tabel 6 Zelfregulatie  $T_1$ 

Item $T_1$	Infit	Outfit	Measure	Error	PTMEA	Item $T_2$	Miscellaneous
70	1.04	.99	1.26	.08	.57	63	
78	1.07	1.00	.35	.07	.63	69	
73	1.15	1.12	.03	.07	.50	64	
63	.89	.90	-.28	.07	.59	62	
71	.96	.95	-.31	.07	.62	71	
76	.91	.93	-1.04	.07	.55	65	
All items							
Mean	1.00	.98	.00	.07		Person Reliability	.54
SD	.09	.07	.70	.00		Person Ability	1.77
All persons						Item Reliability	.99
Mean	.99	.98	-.29	.78		Item Difficulty	13.59
SD	.55	.58	1.17	.13		Cronbach's alpha	.65
Average measures	3	4	5				
	-1.27	-.23	1.02				
Step calibration measures		-1.03	1.03				

De data passen nu redelijk in het Raschmodel, met gemiddelde algemene Infit- en Outfit statistieken van .99 en .98 voor de items en .99 en .98 voor de personen. De gemiddelde maten (average measures) en step calibrations blijken in oplopende volgorde. De schaal heeft een itemmoeilijkheid en itembetrouwbaarheid van 13.59 en .99, en een persoonvaardigheid en persoonsbetrouwbaarheid van 1.77 en .54. De schaal identificeert 13.59 verschillende itemmoeilijkheidsstrata, die onderscheiden worden door de personen en 1.77 verschillende persoonsbekwaamheidsstrata, die door de items worden onderscheiden. De itemmoeilijkheidsstrata zijn ruim boven het vereiste van twee strata, dat minimaal nodig is om een zinvol onderscheid te kunnen maken tussen gemakkelijke of moeilijke items. Een persoonvaardigheidindex van ongeveer 1 in combinatie met een persoonsbetrouwbaarheid rond de .50 betekent dat de meeste personen zich rondom het centrum van de verdeling bevinden, maar dat er nog steeds van uit gegaan kan worden dat de personen aan de bovenkant van de distributie de hoge presteerders zijn en de personen aan de onderkant ervan de zeer matige presteerders (Linacre, 2006, mondelinge mededeling). Volgens de strikte eisen van de persoonsbekwaamheidsstrata zouden 1.77 strata ontoereikend zijn voor het maken van een zinvol onderscheid tussen op het gebied van Zelfregulatie bekwaame of minder bekwaame personen. De range van item calibraties strekt zich uit van -1.04 (item 76) tot 1.26 (item 70) logits. De geordende oplopende rangschikking van de items is enige indicatie voor constructvaliditeit van de schaal, want de items zijn logisch gerangschikt van gemakkelijk tot moeilijk.

De gemiddelde item maat is .00 (SD .70) en de gemiddelde persoonsmaat is -.29 (SD 1.17). Het moeilijkste Zelfregulatie-item is item 70, met 1.26 logits. De gemiddelde item moeilijkheid en gemiddelde persoonsmaat bevinden zich redelijk op eenzelfde niveau. Dit maakt duidelijk dat de itemschaal zowel iets aan de onder-, bovenkant als het midden van het continuüm in deze geoperationaliseerde vorm van Zelfregulatie meet, maar de onderlinge afstanden tussen de items zijn groot (een beter

construct zou dan ook uit meer items moeten bestaan). Daarnaast doen onverwachte antwoordpatronen afbreuk aan de meetkwaliteit (In het crossectionele onderzoek waren er dan ook extra andere items nodig om het construct op een enigszins voor Rasch acceptabele minimum persoonsbetrouwbaarheid te krijgen<sup>1</sup>).

De schaal had een overeenkomstige fit op  $T_2$ , maar het bleek toen niet nodig om de categorieën samen te voegen.

### *Statistiekennis*

Voor het vergelijken van de leeruitkomsten tussen  $T_1$  en  $T_2$  was slechts één SOLO-opgave beschikbaar: een conceptuele open vraag over de betekenis en samenhang van de begrippen gemiddelde, standaarddeviatie en normaalverdeling. Deze opgave is zowel op  $T_1$  als op  $T_2$  volgens de SOLO-taxonomische indeling gecodeerd in vijf categorieën. Zowel de  $T_2$ -opgave als de  $T_1$ -opgave is vervolgens geankerend aan de Rasch itemmoeilijkheid van de opgave in de zeven items tellende SOLO-schaal op  $T_2$ . Voor deze summier vergelijking van de leeruitkomsten is het raadzaam om de volgende aspecten te overwegen:

a) De SOLO-opgave over de vuistregel voor de standaardafwijking bleek in het crossectionele onderzoek (hoofdstuk 5 van dit proefschrift) het op één gemakkelijkste item in de itemhiërarchie van SOLO.

b) Doordat de longitudinale vergelijking beperkt is tot slechts één opgave uit de SOLO-test, kan geen conclusies getrokken worden over andersoortige en vooral moeilijker items uit de SOLO-test en moet hier noodzakelijkerwijs beperkt worden tot het vergelijken van 'retentie', datgene wat is onthouden over statistische concepten. Dit onderzoek pretendeert dan ook geenszins met deze opgave 'kennis en begrip van statistiek' in haar volle omvang te meten en te vergelijken in de tijd. Wel kunnen zonder meetbias eventuele individuele en groepsverschillen in de tijd in persoonsvaardigheid op het gebied van deze specifieke conceptuele kennis getoetst worden.

Omdat de items uit het  $T_1$ -bestand als baseline hebben gefungeerd voor de schaalconstructies, zijn hiervoor de resultaten van de  $T_1$ -Raschanalyses weergegeven. Het uiteindelijk geconstrueerde  $T_2$ -equivalent daarvan en geankerde schalen en maten zijn weergegeven in Appendix A. De persoons- en itemmappen (de grafieken van de persoons- en itemverdelingen binnen de schalen) zijn weergegeven in Appendix B (zowel de 'originele' ongeankerde  $T_1$ -Raschschaal als de op  $T_2$  geankerde  $T_1$ -Rasch-schalen).

1 Bij meerdere exploraties in Rasch blijken vooral 'bredere en meer omvattende' constructen, zoals Zelfregulatie, maar bijvoorbeeld ook Elaboreren en Organiseren, moeilijk voldoende persoonsbetrouwbaarheid te genereren. Tevens blijken in dergelijke constructen negatief geformuleerde items eerder niet dan wel te passen in het model, tenzij er een balans is tussen het aantal positief en negatief geformuleerde items. Bij sommige constructen blijken negatief geformuleerde items daadwerkelijk een ander onderliggend continuüm te meten. Mede doordat 'brede' constructen een problematische meetkwaliteit hebben is Pintrich (2004) er een voorstander van om fijnmaziger te meten. Rasch bevestigt die bevinding.

**Appendix A  $T_2$ -Rasch schalen en ankermeasures****Tabel A.1 Taakwaardering  $T_2$** 

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
14	.98	1.01	1.09	.08	.75	
13	1.06	1.14	1.06	.08	.77	
12	1.06	.99	-.39	.09	.72	
10	.85	.78	-.44	.09	.78	
9	1.09	1.02	-.50	.09	.69	
8	.98	.98	-.82	.09	.68	
All items						
Mean	1.00	.99	.00	.09		Person Reliability .78
SD	.08	.11	.77	.01		Person Ability 2.85
All persons						Item Reliability .99
Mean	.99	.99	1.81	.87		Item Difficulty 11.93
SD	.92	.94	1.94	.23		Cronbach's alpha .85
Average measures	1	2	3	4	5	
	-2.47	-1.01	.23	2.15	4.39	
Step calibration measures		-3.33	-1.33	-.05	4.60	

**Tabel A.2 Zelfeffectiviteit  $T_2$** 

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
21	.81	.79	1.82	.08	.83	
23	1.03	1.09	1.62	.08	.78	
20	.98	.98	.23	.09	.82	
19	.92	.92	-.01	.09	.83	
15	1.42	1.43	-.68	.09	.78	
22	.78	.76	-.73	.09	.85	
26	.99	.94	-2.24	.09	.76	
All items						Person Reliability .88
Mean	.99	.99	.00	.09		Person Ability 3.97
SD	.19	.21	1.31	.01		Item Reliability 1.00
All persons						Item Difficulty 14.73
Mean	.99	.99	.32	.79		Cronbach's alpha .91
SD	.86	.88	2.37	.18		
Average measures	1	2	3	4	5	
	-4.86	-2.27	-.09	2.34	4.60	
Step calibration measures		-4.77	-1.87	1.18	5.45	

Tabel A.3 Testangst T<sub>2</sub>

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
31	.29	.96	.80	.09	.85	
32	.27	.96	.54	.09	.85	
75	.28	.75	.12	.09	.90	
33	.30	1.23	-1.47	.09	.86	
All items						Person Reliability .80
Mean	.98	1.01	.00	.09		Person Ability 3.01
SD	.17	.21	.88	.00		Item Reliability .99
All persons						Item Difficulty 12.76
Mean	1.00	1.01	-1.59	1.08		Cronbach's alpha .90
SD	1.22	1.26	2.55	.35		
Average measures	1	2	3	4	5	
	-5.11	-2.51	-.30	1.63	4.23	
Step calibration measures		-4.96	-.64	1.02	4.59	

Tabel A.4 Leerstrategieën T<sub>2</sub>

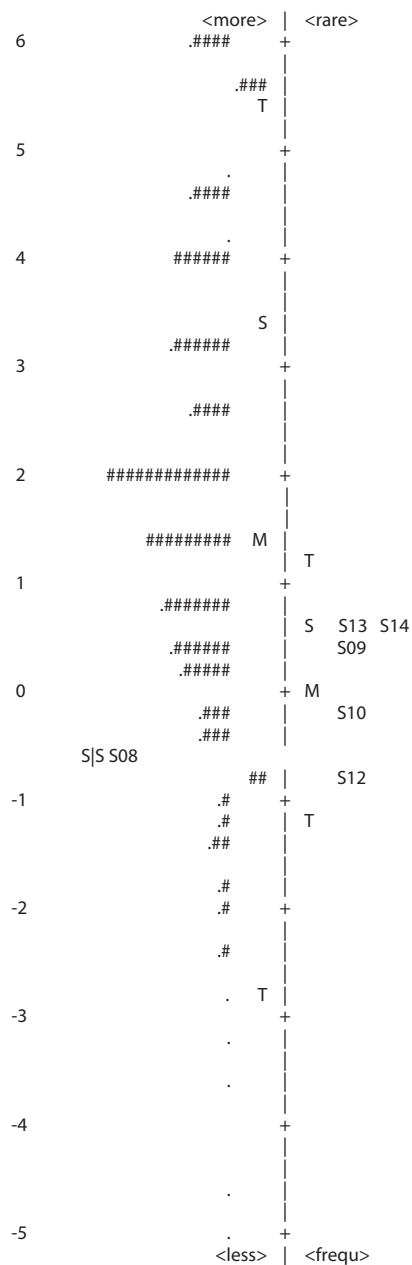
<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>
33	1.28	1.48	.86	.06	.53	33 herhalen
47	1.22	1.38	.61	.06	.47	47 organiseren
49	.86	.87	.24	.07	.50	49 organiseren
34	.77	.70	.16	.07	.62	34 elaboreren
37	.88	.79	.03	.08	.59	37 elaboreren
38	.88	.86	-.85	.09	.48	38 elaboreren
35	.77	.67	-1.05	.10	.58	35 elaboreren
All items						
Mean	.95	.96	.00	.08		Person Reliability .55
SD	.20	.30	.65	.01		Person Ability 1.80
All persons						Item Reliability .99
Mean	1.00	.96	1.05	.73		Item Difficulty 11.24
SD	1.03	1.02	1.16	.26		Cronbach's alpha .62
Average measures	1	2	3	4	5	
	-.85	-.42	-.27	1.13	2.77	
Step calibration measures		-1.84	-.99	-.82	3.65	

Tabel A.5 Zelfregulatie  $T_2$ 

<i>Item</i>	<i>Infit</i>	<i>Outfit</i>	<i>Measure</i>	<i>Error</i>	<i>PTMEA</i>	<i>Miscellaneous</i>	
63	.87	.90	.90	.06	.58	63	
69	1.18	1.23	1.23	.06	.57	69	
62	.88	.89	.89	.06	.58	64	
64	.98	.98	.98	.06	.50	62	
71	1.13	1.09	1.09	.06	.60	71	
65	.93	.98	.98	.07	.40	65	
All items							
Mean	1.00	1.01	1.01	.07		Person Reliability	.54
SD	.12	.12	.12	.01		Person Ability	1.77
All persons						Item Reliability	.99
Mean	1.01	1.01	1.01	.64		Item Difficulty	12.16
SD	.81	.85	.85	.18		Cronbach's alpha	.56
Average measures	1	2	3	4	5		
	-1.03	-.42	.21	.95	2.08		
Step calibration measures		-2.68	-.41	-.21	3.29		

## Appendix B Persoons- en Itemmappen

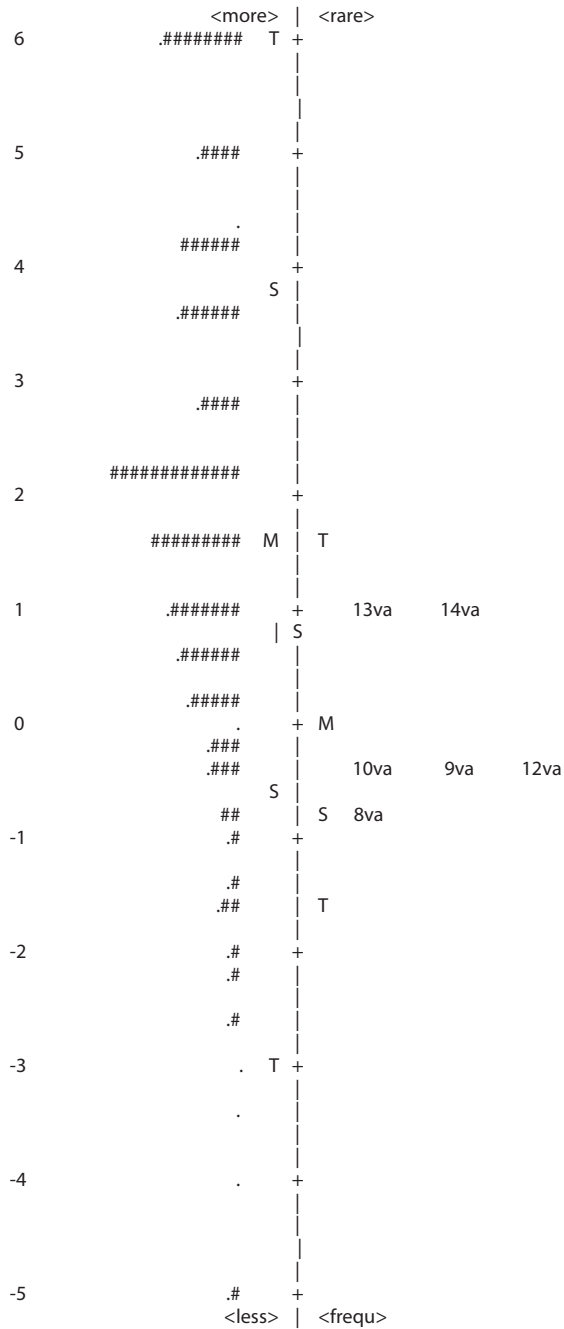
N.B. Per schaal zijn er twee mappen: 1.  $T_1$  ongeankerd; 2.  $T_1$  geankerd.



EACH '#' IS 7.

Figuur 1a Ongeankerde  $T_1$ -Taakwaardering<sup>2</sup>

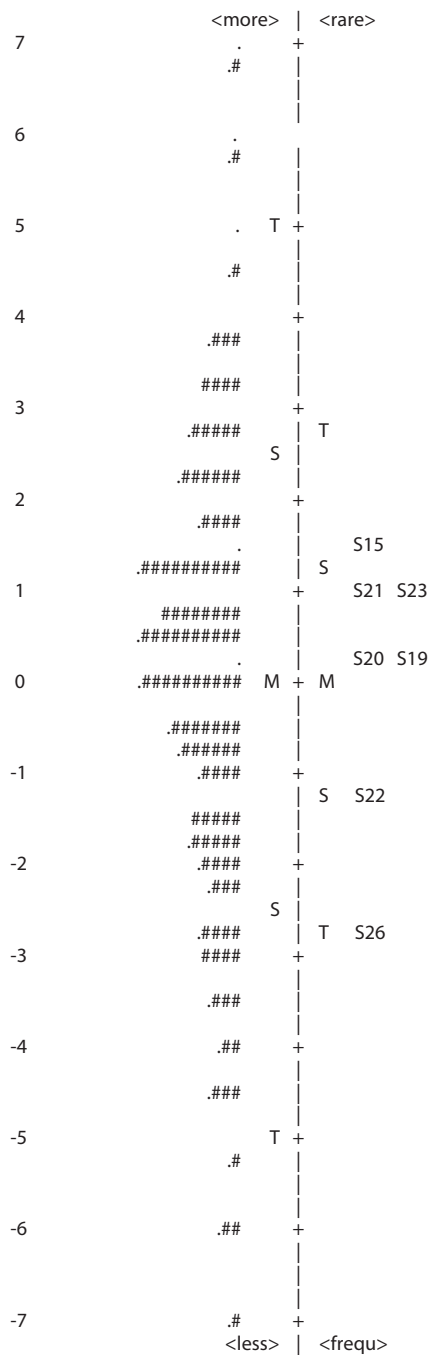
2 Voor het vergemakkelijken van een vergelijking van items in de  $T_2$ -zijn de  $T_1$ -itemnummers gelijk gesteld aan de  $T_2$  item nummers. Dit geldt voor alle hierna volgende  $T_1$  persoons- en itemmappen.



EACH '#' IS 7.

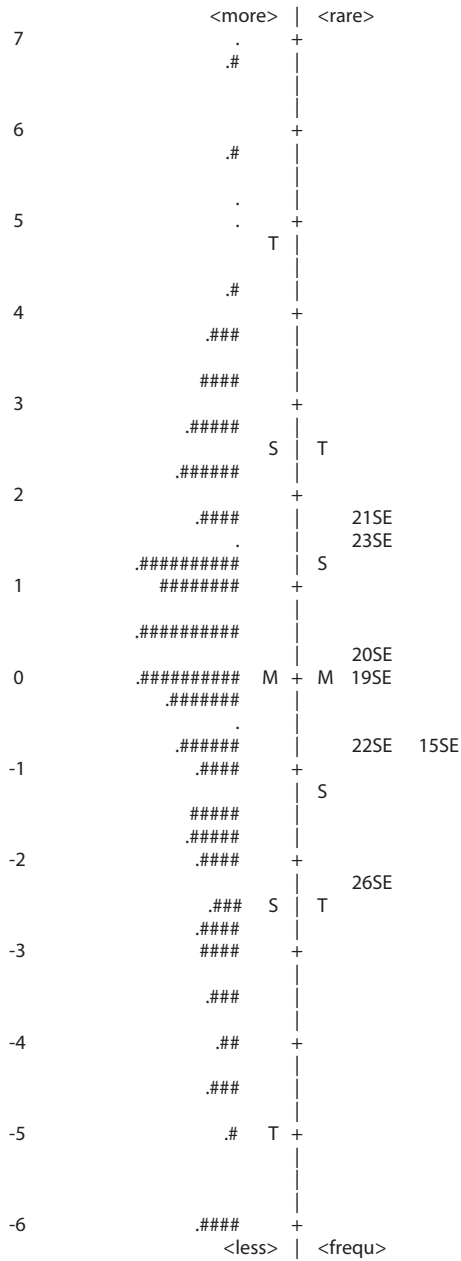
Figuur 1b Geankerde  $T_1$  Taakwaardering





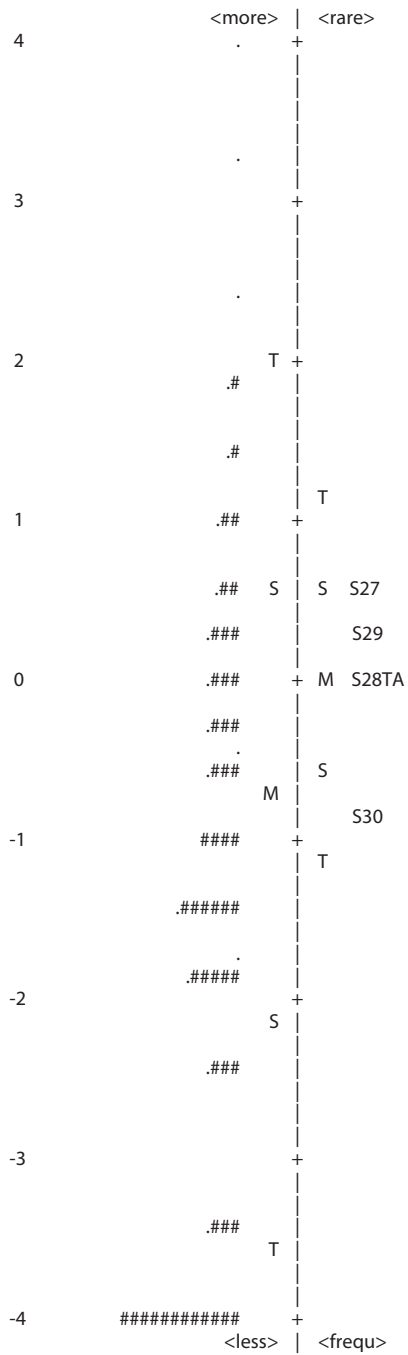
EACH '#' IS 5.

Figuur 2a Ongeankerde Zelfeffectiviteit  $T_1$



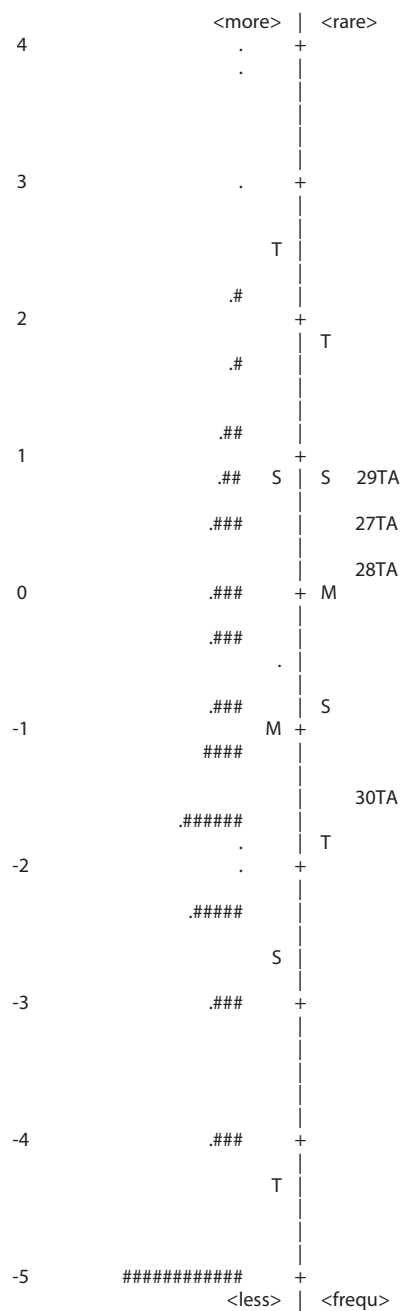
EACH '#' IS 5.

 Figuur 2b Geankerde Zelfeffectiviteit  $T_1$



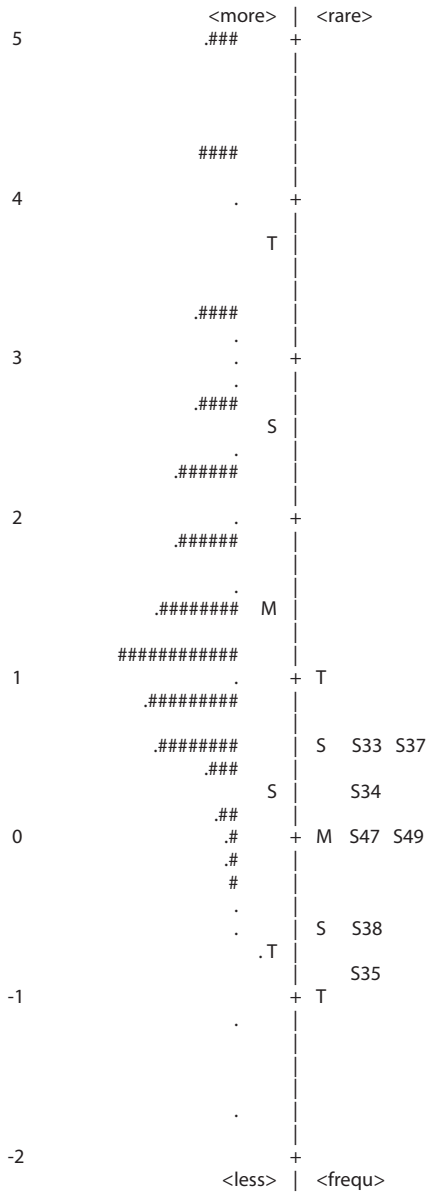
EACH '#' IS 11.

Figuur 3a Ongeankerde Testangst  $T_1$



EACH '#' IS 11.

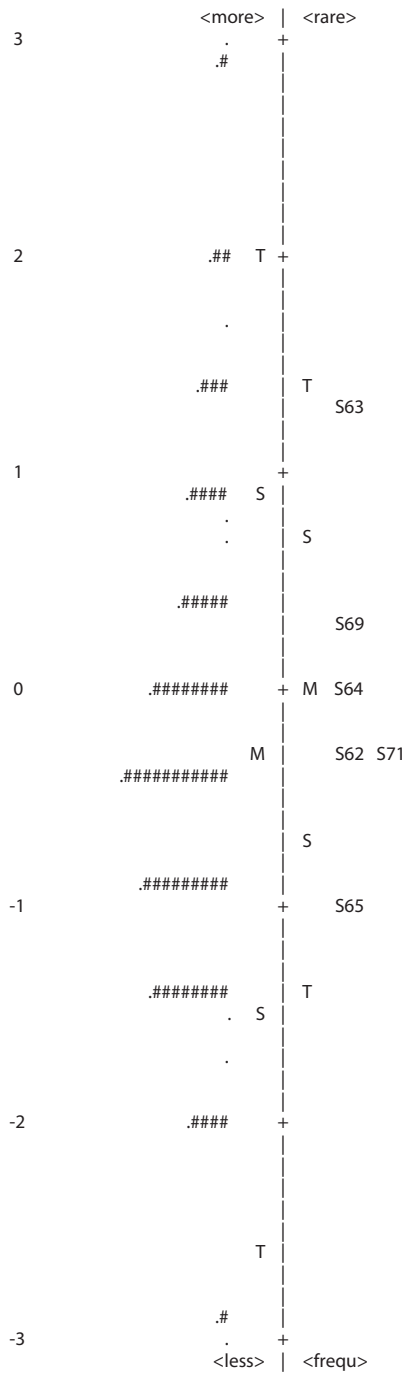
Figuur 3b Geankerde Testangst  $T_1$



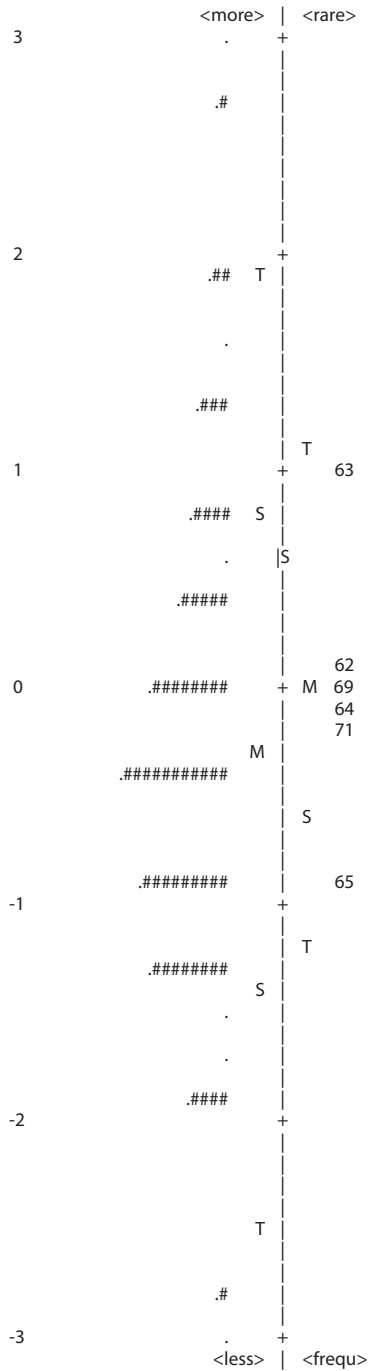
EACH '#' IS 8.

Figuur 4a Ongeankerde Leerstrategieën T<sub>1</sub>





Figuur 5a Ongeankerde Zelfregulatie  $T_1$



EACH '#' IS 10.

Figuur 5b  $T_1$  geankerde Zelfregulatie



